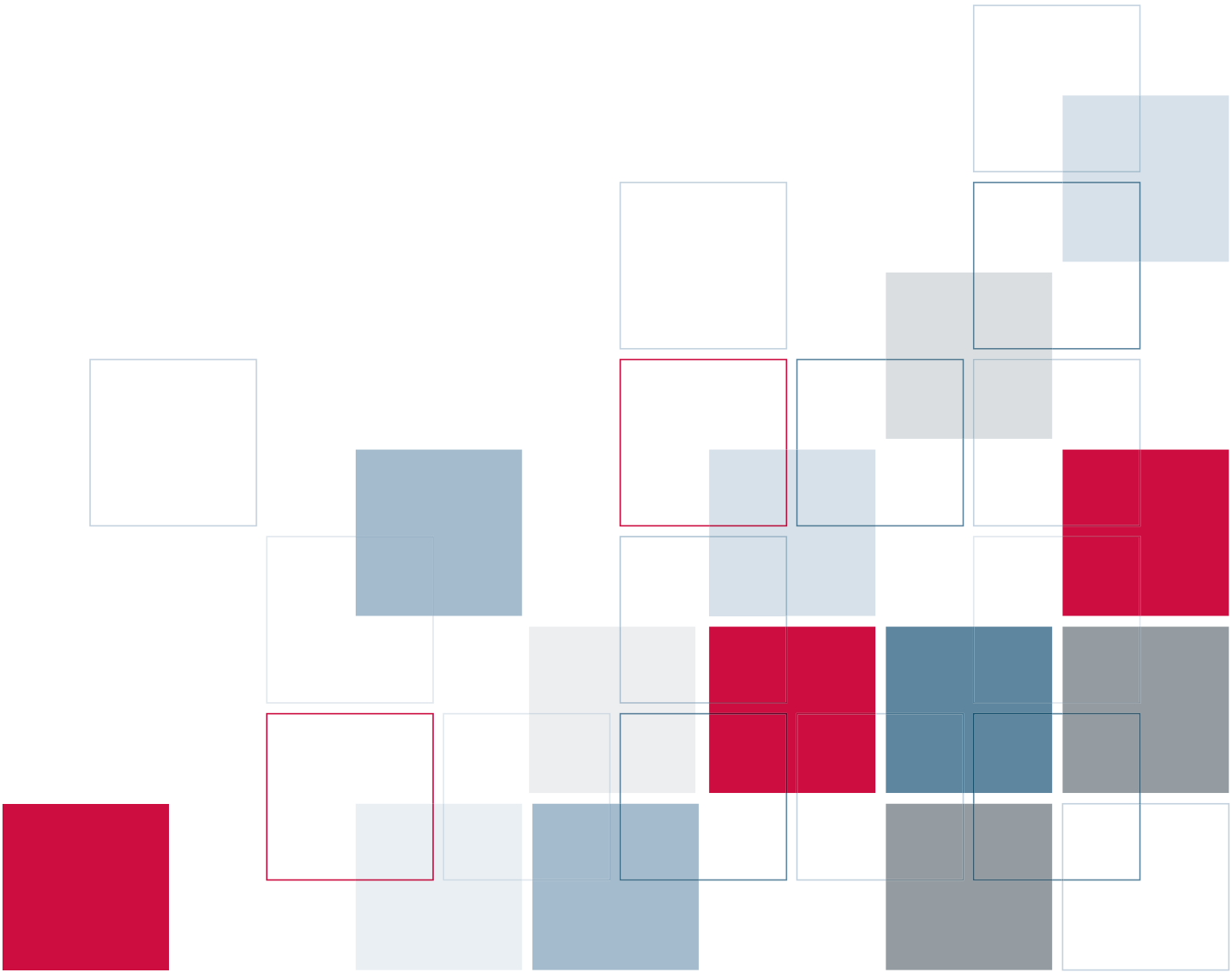
SPSS Complex Samples™ 16.0



Weitere Informationen zu SPSS®-Software-Produkten finden Sie auf unserer Website unter der Adresse *http://www.spss.com* oder wenden Sie sich an

SPSS Inc.
233 South Wacker Drive, 11th Floor
Chicago, IL 60606-6412, USA
Tel.: (312) 651-3000
Fax: (312) 651-3668

SPSS ist eine eingetragene Marke, und weitere Produktnamen sind Marken der SPSS Inc. für Computerprogramme von SPSS Inc. Die Herstellung oder Verbreitung von Materialien, die diese Programme beschreiben, ist ohne die schriftliche Erlaubnis des Eigentümers der Marke und der Lizenzrechte der Software und der Copyrights der veröffentlichten Materialien verboten.

Die SOFTWARE und die Dokumentation werden mit BESCHRÄNKTEN RECHTEN zur Verfügung gestellt. Verwendung, Vervielfältigung und Veröffentlichung durch die Regierung unterliegen den Beschränkungen in Unterabschnitt (c)(1)(ii) von The Rights in Technical Data and Computer Software unter 52.227-7013. Vertragspartner/Hersteller ist SPSS Inc., 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.
Patentnr. 7.023.453

Allgemeiner Hinweis: Andere in diesem Dokument verwendete Produktnamen werden nur zu Identifikationszwecken genannt und können Marken der entsprechenden Unternehmen sein.

Windows ist eine eingetragene Marke der Microsoft Corporation.

Apple, Mac und das Mac-Logo sind Marken von Apple Computer, Inc., die in den USA und in anderen Ländern eingetragen sind.

Dieses Produkt verwendet WinWrap Basic, Copyright 1993–2007, Polar Engineering and Consulting, *http://www.winwrap.com*.

SPSS Complex Samples™ 16.0
Copyright © 2007 SPSS Inc.
Alle Rechte vorbehalten.

Ohne schriftliche Erlaubnis der SPSS GmbH Software darf kein Teil dieses Handbuchs für irgendwelche Zwecke oder in irgendeiner Form mit irgendwelchen Mitteln, elektronisch oder mechanisch, mittels Fotokopie, durch Aufzeichnung oder durch andere Informationsspeicherungssysteme reproduziert werden.

1 2 3 4 5 6 7 8 9 0 10 09 08 07

Vorwort

SPSS 16.0 ist ein umfassendes System zum Analysieren von Daten. Das optionale Erweiterungsmodul SPSS Complex Samples (Komplexe Stichproben) bietet die zusätzlichen Analyseverfahren, die in diesem Handbuch beschrieben sind. Die Prozeduren im Erweiterungsmodul Complex Samples (Komplexe Stichproben) müssen zusammen mit SPSS 16.0 Base verwendet werden. Sie sind vollständig in dieses System integriert.

Installation

Zur Installation von SPSS Complex Samples (Komplexe Stichproben) Erweiterungsmodul führen Sie den Lizenzautorisierungsassistenten mit dem Autorisierungscode aus, den Sie von SPSS erhalten haben. Weitere Informationen finden Sie in den Installationsanweisungen im Lieferumfang von SPSS Complex Samples (Komplexe Stichproben) Erweiterungsmodul.

Kompatibilität

SPSS kann auf vielen Computersystemen ausgeführt werden. Mindestanforderungen an das System und Empfehlungen finden Sie in den Unterlagen, die mit Ihrem System geliefert werden.

Seriennummern

Die Seriennummer des Programms dient gleichzeitig als Identifikationsnummer bei SPSS. Sie benötigen diese Seriennummer, wenn Sie sich an SPSS wenden, um Informationen über Kundendienst, zu Zahlungen oder Aktualisierungen des Systems zu erhalten. Die Seriennummer wird mit dem Base-System ausgeliefert.

Kundendienst

Wenden Sie sich mit Fragen bezüglich der Lieferung oder Ihres Kundenkontos an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden. Halten Sie bitte stets Ihre Seriennummer bereit.

Ausbildungsseminare

SPSS bietet öffentliche und unternehmensinterne Seminare an. Alle Seminare beinhalten auch praktische Übungen. Seminare finden in größeren Städten regelmäßig statt. Wenn Sie weitere Informationen zu diesen Schulungen wünschen, wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden.

Technischer Support

Kunden von SPSS mit Wartungsvertrag können den Technischen Support in Anspruch nehmen. Kunden können sich an den Technischen Support wenden, wenn sie Hilfe bei der Arbeit mit SPSS oder bei der Installation in einer der unterstützten Hardware-Umgebungen benötigen. Informationen über den Technischen Support finden Sie auf der Website von SPSS unter *http://www.spss.com* oder wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden. Bei einem Anruf werden Sie nach Ihrem Namen, dem Namen Ihrer Organisation und Ihrer Seriennummer gefragt.

Weitere Veröffentlichungen

Weitere Exemplare von Produkthandbüchern können direkt bei SPSS Inc. bestellt werden. Besuchen Sie den SPSS Web Store unter *http://www.spss.com/estore* oder wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden. Wenden Sie sich bei telefonischen Bestellungen in den USA und Kanada unter 800-543-2185 direkt an SPSS Inc. Wenden Sie sich bei telefonischen Bestellungen außerhalb von Nordamerika an Ihr regionales SPSS-Büro, das Sie auf der SPPS-Website finden.

Das Handbuch *SPSS Statistical Procedures Companion* von Marija Norušis wurde von Prentice Hall veröffentlicht. Eine neue Fassung dieses Buchs mit Aktualisierungen für SPSS 16.0 ist geplant. Das Handbuch *SPSS Advanced Statistical Procedures Companion*, bei dem auch SPSS 16.0 berücksichtigt wird, erscheint demnächst. Das Handbuch *SPSS Guide to Data Analysis* für SPSS 16.0 wird ebenfalls derzeit erstellt. Ankündigungen für Veröffentlichungen, die ausschließlich über Prentice Hall verfügbar sind, finden Sie auf der SPSS-Website unter *http://www.spss.com/estore* (wählen Sie Ihr Land aus und klicken Sie auf Books).

Kundenmeinungen

Ihre Meinung ist uns wichtig. Teilen Sie uns bitte Ihre Erfahrungen mit SPSS-Produkten mit. Insbesondere haben wir Interesse an neuen, interessanten Anwendungsgebieten von SPSS Complex Samples (Komplexe Stichproben) Erweiterungsmodul. Senden Sie uns eine E-Mail an *suggest@spss.com* oder schreiben Sie an: SPSS Inc., Attn: Director of Product Planning, 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.

Über dieses Handbuch

In diesem Handbuch wird die grafische Benutzeroberfläche für die in SPSS Complex Samples (Komplexe Stichproben) Erweiterungsmodul enthaltenen Prozeduren erläutert. Die Abbildungen der Dialogfelder stammen aus SPSS. Detaillierte Informationen zur Befehlssyntax für die Funktionen in SPSS Complex Samples (Komplexe Stichproben) Erweiterungsmodul sind auf zwei Arten verfügbar: als Bestandteil des umfassenden Hilfesystems und als separates Dokument im PDF-Format im Handbuch *SPSS 16.0 Command Syntax Reference*, das auch über das Menü “Hilfe” verfügbar ist.

Kontakt zu SPSS

Wenn Sie in unseren Verteiler aufgenommen werden möchten, wenden Sie sich an eines unserer Büros, die Sie auf unserer Website unter *http://www.spss.com/worldwide* finden.

Inhalt

Teil I: Benutzerhandbuch

1 Einführung in die Prozeduren von Complex Samples 1

Eigenschaften komplexer Stichproben 1
Verwendung der Prozeduren für komplexe Stichproben 2
Plandateien 3
Weiterführende Literatur 3

2 Stichprobenziehung mithilfe eines komplexen Plans 4

Erstellen eines neuen Stichprobenplans 4
Stichprobenassistent: Stichproben-Variablen 6
Baumsteuerungen zur Navigation im Stichprobenassistenten 7
Stichprobenassistent: Methode der Stichprobenziehung 8
Stichprobenassistent: Stichprobenumfang 10
Ungleiche Umfänge definieren 11
Stichprobenassistent: Ausgabevariablen 12
Stichprobenassistent: Planübersicht 13
Stichprobenassistent: Stichprobe ziehen: Auswahloptionen 14
Stichprobenassistent: Stichprobe ziehen: Ausgabedateien 15
Stichprobenassistent: Fertig stellen 16
Bearbeiten eines bestehenden Stichprobenplans 16
Stichprobenassistent: Planübersicht 17
Ausführen eines bestehenden Stichprobenplans 18
Zusätzliche Funktionen bei den Befehlen CSPLAN und CSSELECT 18

3 Vorbereiten einer komplexen Stichprobe für die Analyse 19

Erstellen eines neuen Analyseplans 20
Analysevorbereitungsassistent: Stichproben-Variablen 21
Baumsteuerungen zur Navigation im Analyseassistenten 22

Analysevorbereitungsassistent: Schätzmethode 22
Analysevorbereitungsassistent: Umfang 23
Ungleiche Umfänge definieren 24
Analysevorbereitungsassistent: Planübersicht 25
Analysevorbereitungsassistent: Fertig stellen 26
Bearbeiten eines bestehenden Analyseplans 26
Analysevorbereitungsassistent: Planübersicht 27

4 Plan für komplexe Stichproben 28

5 Häufigkeiten für komplexe Stichproben 29

Häufigkeiten für komplexe Stichproben: Statistiken 30
Komplexe Stichproben: Fehlende Werte 31
Komplexe Stichproben: Optionen 32

6 Deskriptive Statistiken für komplexe Stichproben 33

Deskriptive Statistiken für komplexe Stichproben: Statistiken 34
Deskriptive Statistiken für komplexe Stichproben: Fehlende Werte 35
Komplexe Stichproben: Optionen 36

7 Kreuztabellen für komplexe Stichproben 37

Kreuztabellen für komplexe Stichproben - Statistik 39
Komplexe Stichproben: Fehlende Werte 40
Komplexe Stichproben: Optionen 41

8 Verhältnisse für komplexe Stichproben 42

Verhältnisse für komplexe Stichproben: Statistiken 43
Verhältnisse für komplexe Stichproben: Fehlende Werte 44
Komplexe Stichproben: Optionen 45

9 Allgemeines lineares Modell für komplexe Stichproben 46

Allgemeines lineares Modell für komplexe Stichproben: Statistik 49
Hypothesentests für komplexe Stichproben 50
Allgemeines lineares Modell für komplexe Stichproben: Geschätzte Mittelwerte 52
Allgemeines lineares Modell für komplexe Stichproben: Speichern 53
Allgemeines lineares Modell für komplexe Stichproben: Optionen 54
Zusätzliche Funktionen beim Befehl CSGLM 54

10 Logistische Regression für komplexe Stichproben 56

Logistische Regression für komplexe Stichproben: Referenzkategorie 57
Logistische Regression für komplexe Stichproben: Modell 58
Logistische Regression für komplexe Stichproben: Statistik 60
Hypothesentests für komplexe Stichproben 61
Logistische Regression für komplexe Stichproben: Quotenverhältnis 62
Logistische Regression für komplexe Stichproben: Speichern 63
Logistische Regression für komplexe Stichproben: Optionen 64
Zusätzliche Funktionen beim Befehl CSLOGISTIC 65

11 Ordinale Regression für komplexe Stichproben 66

Ordinale Regression für komplexe Stichproben: Antwortwahrscheinlichkeiten 68
Ordinale Regression für komplexe Stichproben: Modell 69
Ordinale Regression für komplexe Stichproben: Statistik 70
Hypothesentests für komplexe Stichproben 72
Ordinale Regression für komplexe Stichproben: Quotenverhältnisse 73
Ordinale Regression für komplexe Stichproben: Speichern 74
Ordinale Regression für komplexe Stichproben: Optionen 75
Zusätzliche Funktionen beim Befehl CSORDINAL 76

12 Cox-Regression für komplexe Stichproben 78

Ereignis definieren 81

Einflussvariablen 82
Definieren einer zeitabhängigen Einflussvariablen 83
Untergruppen 85
Modell 86
Statistik 88
Diagramme 90
Hypothesentests 91
Speichern 92
Exportieren 94
Optionen 96
Zusätzliche Funktionen des CSCOXREG-Befehls 98

Teil II: Beispiele

13 Stichprobenassistent für komplexe Stichproben 100

Ziehen einer Stichprobe aus einem vollständigen Stichprobenrahmen 100
Verwendung des Assistenten 100
Planübersicht 110
Stichprobenübersicht 110
Stichprobenergebnisse 111
Ziehen einer Stichprobe aus einem partiellen Stichprobenrahmen 112
Verwenden des Assistenten für die Stichprobenziehung aus dem ersten Teilrahmen 112
Stichprobenergebnisse 125
Verwenden des Assistenten für die Stichprobenziehung aus dem zweiten Teilrahmen 125
Stichprobenergebnisse 130
Stichprobenziehung mit PPS (Probability Proportional to Size; Wahrscheinlichkeit proportional zur Größe) 130
Verwendung des Assistenten 131
Planübersicht 142
Stichprobenübersicht 143
Stichprobenergebnisse 145
Verwandte Prozeduren 147

14 Analysevorbereitungsassistent für komplexe Stichproben 148

Verwendung des Analysevorbereitungsassistenten für komplexe Stichproben zur Vorbereitung von öffentlich zugänglichen NHIS-Daten . . . 148
Verwendung des Assistenten . . . 148
Auswertung . . . 151
Vorbereitung für die Analyse, wenn die Datendatei keine Stichprobengewichte enthält . . . 151
Berechnung von Einschlusswahrscheinlichkeiten und Stichprobengewichten . . . 151
Verwendung des Assistenten . . . 154
Auswertung . . . 162
Verwandte Prozeduren . . . 162

15 Häufigkeiten für komplexe Stichproben 163

Verwendung von "Häufigkeiten für komplexe Stichproben" zur Analyse der Verwendung von Nahrungsergänzungen . . . 163
Durchführung der Analyse . . . 163
Häufigkeitstabelle . . . 166
Häufigkeit nach Teilgesamtheit . . . 167
Zusammenfassung . . . 167
Verwandte Prozeduren . . . 168

16 Deskriptive Statistiken für komplexe Stichproben 169

Verwendung von "Deskriptive Statistiken für komplexe Stichproben" zur Analyse von Aktivitätsniveaus . . . 169
Durchführung der Analyse . . . 169
Univariate Statistiken . . . 172
Univariate Statistiken nach Teilgesamtheit . . . 172
Zusammenfassung . . . 173
Verwandte Prozeduren . . . 173

17 Kreuztabellen für komplexe Stichproben 174

Verwendung von "Kreuztabellen für komplexe Stichproben" zum Messen des relativen Risikos eines Ereignisses . . . 174
Durchführung der Analyse . . . 174
Kreuztabellen . . . 177

Risikoschätzer . . . 178
Risikoschätzer nach Teilgesamtheit . . . 179
Zusammenfassung . . . 179
Verwandte Prozeduren . . . 180

18 Verhältnisse für komplexe Stichproben 181

Verwenden von "Verhältnisse für komplexe Stichproben" zur Erleichterung der Schätzung von Immobilienwerten . . . 181
Durchführung der Analyse . . . 181
Verhältnisse . . . 183
Pivotierte Verhältnistabelle . . . 184
Zusammenfassung . . . 184
Verwandte Prozeduren . . . 185

19 Allgemeines lineares Modell für komplexe Stichproben 186

Verwendung des allgemeinen linearen Modells für komplexe Stichproben zur Anpassung einer zweifaktoriellen ANOVA . . . 186
Durchführung der Analyse . . . 186
Modellzusammenfassung . . . 191
Tests der Modelleffekte . . . 191
Parameterschätzer . . . 192
Geschätzte Randmittel . . . 193
Zusammenfassung . . . 196
Verwandte Prozeduren . . . 196

20 Logistische Regression für komplexe Stichproben 197

Verwenden der logistischen Regression für komplexe Stichproben zur Bewertung des Kreditrisikos . . . 197
Durchführung der Analyse . . . 197
Werte für Pseudo-R-Quadrat . . . 201
Klassifikation . . . 202
Tests der Modelleffekte . . . 203
Parameterschätzer . . . 203
Quotenverhältnisse . . . 204
Zusammenfassung . . . 206
Verwandte Prozeduren . . . 206

21 Ordinale Regression für komplexe Stichproben 207

Verwendung der ordinalen Regression für komplexe Stichproben zur Analyse von Umfrageergebnissen 207
- Durchführung der Analyse 207
- Werte für Pseudo-R-Quadrat 212
- Tests der Modelleffekte 213
- Parameterschätzer 213
- Klassifikation 215
- Quotenverhältnisse 216
- Verallgemeinertes kumulatives Modell 217
- Verwerfen nichtsignifikanter Einflussvariablen 218
- Warnungen 220
- Vergleichen von Modellen 221
- Zusammenfassung 222

Verwandte Prozeduren 222

22 Cox-Regression für komplexe Stichproben 223

Verwenden einer zeitabhängigen Einflussvariablen in der Cox-Regression für komplexe Stichproben 223
- Vorbereitung der Daten 223
- Durchführung der Analyse 228
- Informationen zum Stichprobenplan 233
- Tests der Modelleffekte 234
- Test für proportionale Hazard-Raten 234
- Hinzufügen einer zeitabhängigen Einflussvariablen 235

Mehrere Fälle pro Subjekt in “Cox-Regression für komplexe Stichproben” 238
- Vorbereiten der Daten für die Analyse 239
- Erstellen eines Analyseplans für einfache Zufallsstichprobenziehungen 254
- Durchführung der Analyse 258
- Informationen zum Stichprobenplan 266
- Tests der Modelleffekte 267
- Parameter-Schätzer 267
- Musterwerte 268
- Log-Minus-Log-Diagramm 269
- Zusammenfassung 269

Anhang

A Beispieldateien *271*

Bibliografie *283*

Index *285*

Teil I: Benutzerhandbuch

Kapitel

1

Einführung in die Prozeduren von Complex Samples

Eine Grundannahme bei analytischen Prozeduren in herkömmlichen Softwarepaketen ist, dass die Beobachtungen in einer Datendatei eine einfache Zufallsstichprobe aus der zu betrachtenden Grundgesamtheit darstellen. Diese Annahme ist für eine wachsende Anzahl von Unternehmen und Wissenschaftler unhaltbar, für die es kostengünstig und zweckmäßig ist, Stichproben auf strukturiertere Weise zu gewinnen.

Mit der Option "Complex Samples" (Komplexe Stichproben) können Sie eine Stichprobe nach einem komplexen Plan auswählen und die Planspezifikationen in die Datenanalyse integrieren, um sicherzustellen, dass die Ergebnisse gültig sind.

Eigenschaften komplexer Stichproben

Eine komplexe Stichprobe kann sich in verschiedener Hinsicht von einer einfachen Zufallsstichprobe unterscheiden. Bei einer einfachen Zufallsstichprobe werden die einzelnen Stichprobeneinheiten zufällig mit gleicher Wahrscheinlichkeit und ohne Zurücklegen (OZ) aus der gesamten Grundgesamtheit ausgewählt. Im Gegensatz dazu kann eine komplexe Stichprobe einige oder alle der folgenden Merkmale aufweisen:

Schichtung. Bei einer geschichteten Stichprobenziehung werden die Stichproben unabhängig voneinander innerhalb von sich nicht überschneidenden Untergruppen der Grundgesamtheit, den so genannten Schichten, ausgewählt. Beispiele für Schichten sind sozioökonomische Gruppen, Berufsgruppen, Altersgruppen oder ethnische Gruppen. Bei Verwendung einer Schichtung können Sie angemessene Stichprobengrößen für zu untersuchende Untergruppen gewährleisten, die Genauigkeit von Gesamtschätzungen verbessern und unterschiedliche Stichprobenverfahren für die verschiedenen Schichten verwenden.

Klumpenbildung. Zur Ziehung von Klumpenstichproben gehört die Auswahl von Gruppen von Stichprobeneinheiten, so genannter Klumpen. Beispiele für Klumpen sind Schulen, Krankenhäuser oder geografische Gebiete; die dazugehörigen Stichprobeneinheiten sind Schüler, Patienten bzw. Einwohner. Klumpenbildung ist bei mehrstufigen Plänen und Gebietsstichproben (geografischen Stichproben) üblich.

Mehrere Stufen. Bei einer mehrstufigen Stichprobenziehung wird zunächst auf der Grundlage von Klumpen eine Stichprobe für die erste Stufe ausgewählt. Dann wird eine Stichprobe der zweiten Stufe ausgewählt, indem aus den ausgewählten Klumpen Teilstichproben gezogen werden. Wenn die Stichprobe der zweiten Stufe auf Teilklumpen beruht, können Sie eine dritte Stufe zur Stichprobe hinzufügen. In der ersten Stufe einer Umfrage könnte beispielsweise eine Stichprobe von Städten gezogen werden. Aus den ausgewählten Städten könnten dann Stichproben der

Haushalte gezogen werden. Schließlich könnten einzelne Personen aus den ausgewählten Haushalten befragt werden. Mit dem Stichproben- und dem Analysevorbereitungsassistenten können Sie drei Stufen in einem Plan angeben.

Ziehen nichtzufälliger Stichproben. Wenn eine zufällige Auswahl schwer zu erzielen ist, können die Stichprobeneinheiten systematisch (in festgelegten Intervallen) oder sequenziell gezogen werden.

Ungleiche Auswahlwahrscheinlichkeiten. Bei der Ziehung von Klumpen, die jeweils eine andere Anzahl von Einheiten enthalten, können Sie eine PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) für die Stichprobenziehung verwenden. Diese gewährleistet, dass die Auswahlwahrscheinlichkeit eines Klumpens dem Anteil an Einheiten entspricht, die er enthält. Bei der PPS-Stichprobenziehung können auch allgemeinere Gewichtungsschemata für die Auswahl der Einheiten verwendet werden.

Unbeschränkte Stichprobenziehung. Bei der unbeschränkten Stichprobenziehung werden Einheiten mit Zurücklegen (MZ) ausgewählt. Eine Einheit kann also mehrmals für die Stichprobe ausgewählt werden.

Stichprobengewichte. Stichprobengewichte werden beim Ziehen komplexer Stichproben automatisch berechnet und entsprechen idealerweise der “Häufigkeit”, die jede Stichprobeneinheit in der Ziel-Grundgesamtheit aufweist. Daher sollte die Summe der Gewichte in der Stichprobe einen Schätzwert für den Umfang der Grundgesamtheit darstellen. Für die Analyseverfahren in “Komplexe Stichproben” sind Stichprobengewichte für die ordnungsgemäße Analyse komplexer Stichproben erforderlich. Hinweis: Diese Gewichte sollten ausschließlich in der Option “Komplexe Stichproben” und nicht bei anderen Analyseverfahren über die Prozedur “Fälle gewichten” verwendet werden. Bei der Prozedur “Fälle gewichten” werden die Gewichte als Fallreplikationen behandelt.

Verwendung der Prozeduren für komplexe Stichproben

Welche Verfahren für komplexe Stichproben für Sie infrage kommen, hängt von Ihren jeweiligen Bedürfnissen ab. Die Hauptbenutzertypen haben folgende Ziele:

- Planung und Durchführung von Studien anhand komplexer Pläne, eventuell spätere Analyse der Stichprobe Das wichtigste Werkzeug für Personen, die Studien durchführen, ist der Stichprobenassistent.
- Analysieren von Dateien mit Stichprobendaten, die zuvor anhand komplexer Pläne gewonnen wurden Bevor Sie die Analyseverfahren für komplexe Stichproben nutzen können, benötigen Sie möglicherweise den Analysevorbereitungsassistenten.

Unabhängig davon, welcher Benutzertyp Sie sind, müssen Sie für die Prozeduren für komplexe Stichproben Planinformationen angeben. Diese Informationen werden zur einfacheren Wiederverwendung in einer **Plandatei** gespeichert.

Plandateien

Eine Plandatei enthält Spezifikationen für komplexe Stichproben. Es gibt zwei Typen von Plandateien:

Stichprobenplan. Durch die im Stichprobenassistenten angegebenen Spezifikationen wird ein Stichprobenplan definiert, der zum Ziehen von komplexen Stichproben verwendet wird. Diese Spezifikationen sind in der Stichprobenplan-Datei enthalten. Eine Stichprobenplan-Datei enthält außerdem einen Standard-Analyseplan, der für den angegebenen Stichprobenplan geeignete Schätzmethoden verwendet.

Analyseplan. Diese Plandatei enthält Informationen, die bei den Analyseverfahren in "Komplexe Stichproben" benötigt werden, um die Varianzschätzungen für komplexe Stichproben ordnungsgemäß zu berechnen. Zum Plan gehören die Stichprobenstruktur, Schätzmethoden für die einzelnen Stufen und Verweise auf erforderliche Variablen, wie beispielsweise die Stichprobengewichte. Mit dem Analysevorbereitungsassistenten können Sie Analysepläne erstellen und bearbeiten.

Das Speichern der Angaben in einer Plandatei bringt verschiedene Vorteile mit sich, unter anderem folgende:

- Personen, die Studien durchführen, können die erste Stufe eines mehrstufigen Stichprobenplans angeben und die Einheiten der ersten Stufe sofort ziehen, Informationen für die Ziehung der Stichprobeneinheiten der zweiten Stufe sammeln und dann die zweite Stufe in den Stichprobenplan integrieren.
- Ein Analytiker, dem die Stichprobenplan-Datei nicht zugänglich ist, kann einen Analyseplan angeben und bei jedem Analyseverfahren für komplexe Stichproben auf diesen Plan zurückgreifen.
- Ein Entwickler großer öffentlich zugänglicher Stichproben (Public-Use-Stichproben) kann die Stichprobenplan-Datei veröffentlichen und damit die Anweisungen für die Analytiker vereinfachen und ermöglichen, dass nicht jeder Analytiker einen eigenen Analyseplan entwickeln muss.

Weiterführende Literatur

Weitere Informationen zu Stichprobenverfahren finden Sie in folgenden Texten:

Cochran, W. G. 1977. *Sampling Techniques*, 3rd (Hg.). New York: John Wiley and Sons.

Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons.

Kish, L. 1987. *Statistical Design for Research*. New York: John Wiley and Sons.

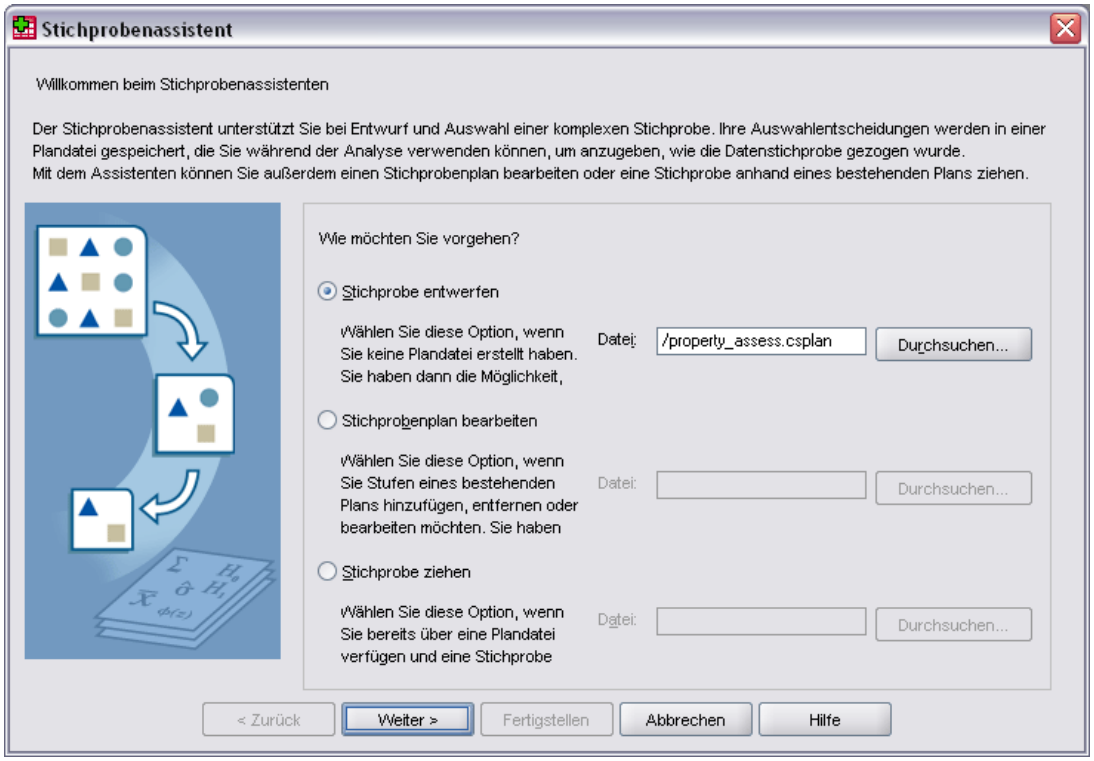
Murthy, M. N. 1967. *Sampling Theory and Methods*. Kalkutta, Indien: Statistical Publishing Society.

Särndal, C., B. Swensson, als auch J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Kapitel
2

Stichprobenziehung mithilfe eines komplexen Plans

Abbildung 2-1
Stichprobenassistent – Schritt "Willkommen"


Stichprobenassistent
Willkommen beim Stichprobenassistenten
Der Stichprobenassistent unterstützt Sie bei Entwurf und Auswahl einer komplexen Stichprobe. Ihre Auswahlentscheidungen werden in einer Plandatei gespeichert, die Sie während der Analyse verwenden können, um anzugeben, wie die Datenstichprobe gezogen wurde.
Mit dem Assistenten können Sie außerdem einen Stichprobenplan bearbeiten oder eine Stichprobe anhand eines bestehenden Plans ziehen.
Wie möchten Sie vorgehen?
Stichprobe entwerfen
Wählen Sie diese Option, wenn Sie keine Plandatei erstellt haben. Sie haben dann die Möglichkeit,
Datei:
/property_assess.csplan
Durchsuchen...
Stichprobenplan bearbeiten
Wählen Sie diese Option, wenn Sie Stufen eines bestehenden Plans hinzufügen, entfernen oder bearbeiten möchten. Sie haben
Datei:
Durchsuchen...
Stichprobe ziehen
Wählen Sie diese Option, wenn Sie bereits über eine Plandatei verfügen und eine Stichprobe
Datei:
Durchsuchen...
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

Der Stichprobenassistent führt Sie durch die Schritte zum Erstellen, Bearbeiten bzw. Ausführen einer Stichprobenplan-Datei. Vor der Verwendung des Assistenten sollten Sie über eine klar umrissene Ziel-Grundgesamtheit und eine Liste der Stichprobeneinheiten verfügen und einen geeigneten Stichprobenplan im Kopf haben.

Erstellen eines neuen Stichprobenplans

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Komplexe Stichproben
    Stichprobe auswählen...

- Wählen Sie die Option Stichprobe entwerfen und wählen Sie einen Dateinamen für die Plandatei, in der der Stichprobenplan gespeichert werden soll.
- Klicken Sie auf Weiter, um unter Verwendung des Assistenten fortzufahren.
- Optional können Sie im Schritt "Stichproben-Variablen" Schichten, Klumpen und Eingabe-Stichprobengewichte definieren. Klicken Sie anschließend auf Weiter.
- Optional können Sie im Schritt "Methode der Stichprobenziehung" eine Methode für die Auswahl der Items auswählen.

  Bei Auswahl von PPS Brewer oder PPS Murthy können Sie auf Fertig stellen klicken, um die Stichprobe zu ziehen. Anderenfalls klicken Sie auf Weiter und gehen Sie dann folgendermaßen vor:

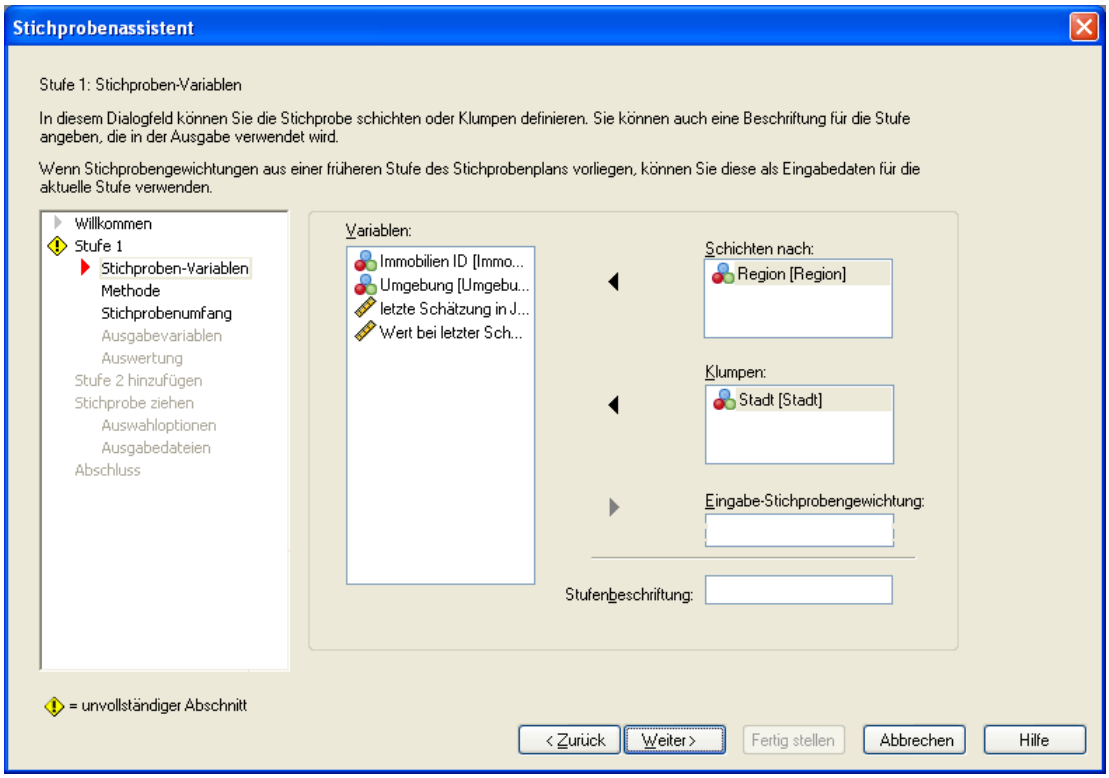
- Geben Sie im Schritt "Stichprobenumfang" die Anzahl bzw. den Anteil der Einheiten für die Stichprobenziehung an.
- Jetzt können Sie auf Fertig stellen klicken, um die Stichprobe zu ziehen.

  Optional können Sie in weiteren Schritten folgende Aktionen durchführen:

  - Ausgabevariablen auswählen, die gespeichert werden sollen.
  - Hinzufügen einer zweiten oder dritten Stufe zum Plan.
  - Festlegen verschiedener Auswahloptionen, u. a. der folgenden: die Angabe, aus welchen Stufen die Stichproben gezogen werden sollen, der Startwert für Zufallszahlen und die Angabe, ob benutzerdefinierte fehlende Werte als gültige Werte von Stichproben-Variablen behandelt werden sollen.
  - Die Auswahl des Speicherorts für die Ausgabedaten.
  - Einfügen der getroffenen Auswahl als Befehlssyntax.

Stichprobenassistent: Stichproben-Variablen

Abbildung 2-2
Stichprobenassistent – Schritt "Stichproben-Variablen"



In diesem Schritt können Sie die Schichtungs- und Klumpenvariablen auswählen und Eingabe-Stichprobengewichte definieren. Außerdem können Sie eine Beschriftung für die Stufe angeben.

Schichten nach. Durch die Kombination von Schichtungsvariablen werden eindeutige Teilgesamtheiten, so genannte Schichten, definiert. Aus jeder Schicht werden getrennte Stichproben gewonnen. Zur Verbesserung der Genauigkeit Ihrer Schätzungen sollten Einheiten innerhalb von Schichten für die zu untersuchenden Merkmale so homogen wie möglich sein.

Klumpen. Klumpenvariablen definieren Gruppen von Beobachtungseinheiten, so genannte Klumpen. Die Definition von Klumpen ist sinnvoll, wenn eine unmittelbare Stichprobenziehung der Beobachtungseinheiten aus der Grundgesamtheit kostspielig oder nicht möglich ist. Stattdessen können Sie Klumpen aus der Grundgesamtheit ziehen und dann aus den ausgewählten Klumpen die Stichprobe der Beobachtungseinheiten ziehen. Die Verwendung von Klumpen kann jedoch zu Korrelationen zwischen den Stichprobeneinheiten führen, wodurch die Genauigkeit verringert wird. Um diese Wirkung möglichst gering zu halten, sollten die Einheiten innerhalb der Klumpen für die zu untersuchenden Merkmale so heterogen wie möglich sein. Für einen mehrstufigen Plan muss mindestens eine Klumpenvariable definiert werden. Klumpen sind außerdem erforderlich, wenn mehrere verschiedene Stichprobenmethoden verwendet werden sollen. Für weitere Informationen siehe Stichprobenassistent: Methode der Stichprobenziehung auf S. 8.

Eingabe-Stichprobengewichtung. Wenn der aktuelle Stichprobenplan Teil eines größeren Stichprobenplans ist, können Stichprobengewichte aus einer früheren Stufe des größeren Plans vorliegen. In der ersten Stufe des aktuellen Plans können Sie eine numerische Variable angeben, die diese Gewichte enthält. Die Stichprobengewichte für die weiteren Stufen des aktuellen Plans werden automatisch berechnet.

Stufenbeschriftung. Sie können für jede Stufe ein optionales String-Label angeben. Dieses wird in der Ausgabe verwendet, um die stufenweisen Informationen besser identifizieren zu können.

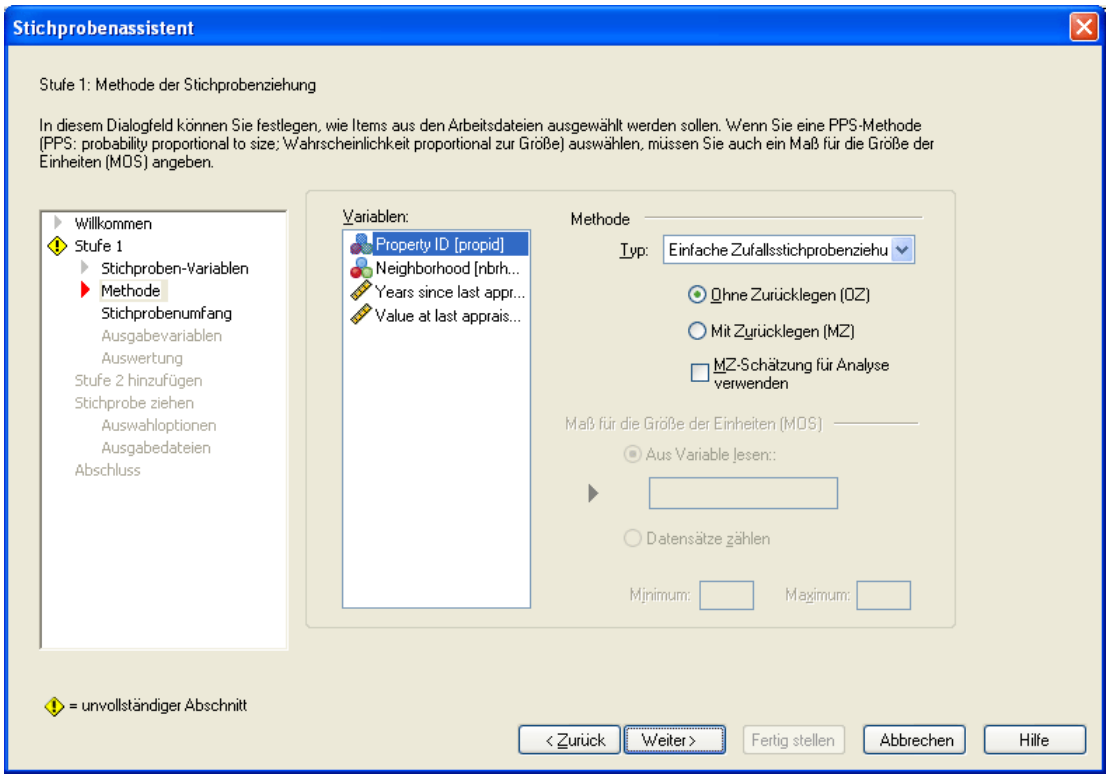
Anmerkung: Die Liste der Quellvariablen hat in allen Schritten des Assistenten denselben Inhalt. Anders ausgedrückt: Variablen, die in einem Schritt aus der Liste der Quellvariablen entfernt werden, werden in allen Schritten aus der Liste entfernt. Variablen, die wieder zur Liste der Quellvariablen hinzugefügt werden, werden in allen Schritten in der Liste angezeigt.

Baumsteuerungen zur Navigation im Stichprobenassistenten

Auf der linken Seite jedes Schritts im Stichprobenassistenten finden Sie eine Gliederung, die eine Übersicht über alle Schritte bietet. Sie können im Assistenten navigieren, indem Sie in der Gliederung auf den Namen eines aktivierten Schrittes klicken. Schritte sind aktiviert, wenn alle vorangegangenen Schritte gültig sind, d. h. wenn für jeden vorangegangen Schritt die erforderlichen Mindestangaben vorgenommen wurden. Weitere Informationen dazu, warum ein Schritt möglicherweise ungültig ist, finden Sie in der Hilfe zu den einzelnen Schritten.

Stichprobenassistent: Methode der Stichprobenziehung

Abbildung 2-3
Stichprobenassistent – Schritt "Methode der Stichprobenziehung"



In diesem Schritt können Sie angeben, wie Fälle aus der Arbeitsdatei ausgewählt werden sollen.

Methode. Die Steuerelemente in dieser Gruppe werden zur Festlegung einer Auswahlmethode verwendet. Bei einigen Arten der Stichprobenziehung können Sie auswählen, ob die Stichprobenziehung mit Zurücklegen (MZ) oder ohne Zurücklegen (OZ) erfolgen soll. Weitere Informationen finden Sie in den Beschreibungen zu den verschiedenen Arten. Hinweis: einige PPS-Methoden (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) sind nur verfügbar, wenn Klumpen definiert wurden, und alle PPS-Methoden sind nur in der ersten Stufe eines Plans verfügbar. Außerdem sind Methoden MZ nur in der letzten Stufe eines Plans verfügbar.

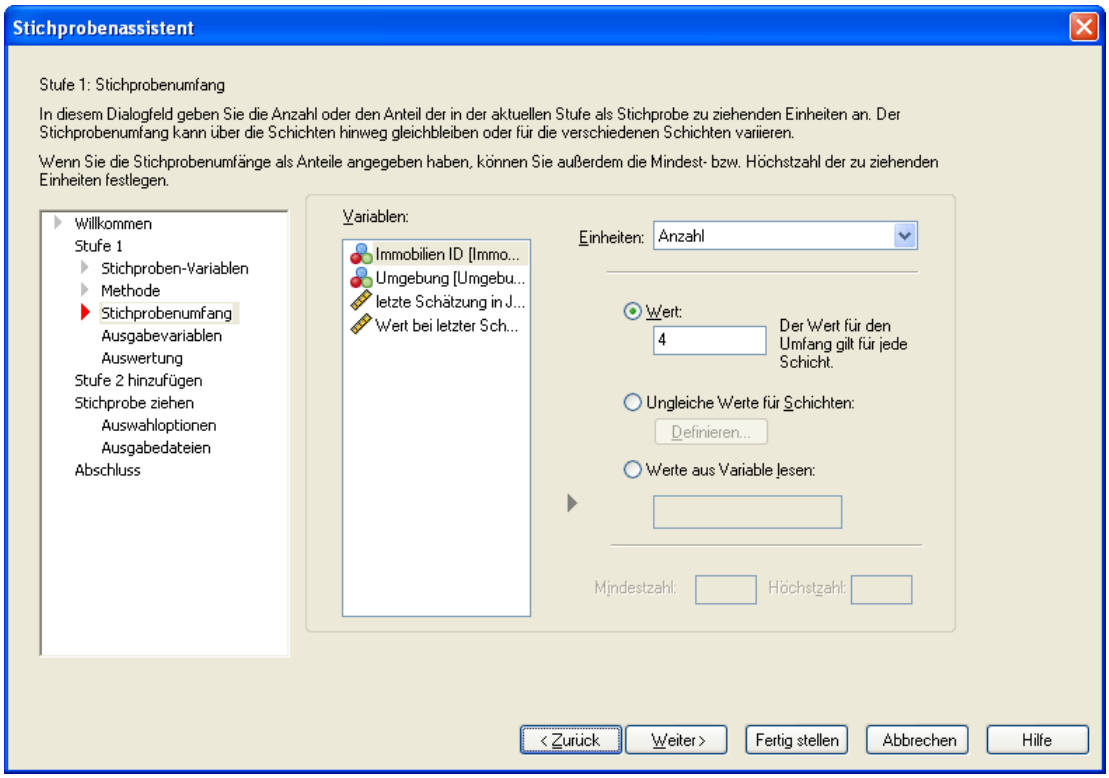
- **Einfache Zufallsstichprobenziehung.** Die Einheiten werden mit gleicher Wahrscheinlichkeit ausgewählt. Sie können mit oder ohne Zurücklegen ausgewählt werden.
- **Einfach systematisch.** Die Einheiten werden in festgelegten Intervallen im gesamten Stichprobenrahmen (bzw. den Schichten, wenn sie angegeben wurden) ausgewählt und ohne Zurücklegen gezogen. Eine zufällig ausgewählte Einheit im ersten Intervall wird als Startpunkt festgelegt.
- **Einfach sequenziell.** Die Einheiten werden sequenziell mit gleicher Wahrscheinlichkeit und ohne Zurücklegen ausgewählt.

- **PPS.** Dies ist eine Methode für die erste Stufe, bei der Einheiten zufällig ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zum Umfang. Alle Einheiten können mit Zurücklegen ausgewählt werden; nur Klumpen können ohne Zurücklegen ausgewählt werden.
- **PPS systematisch.** Dies ist eine Methode für die erste Stufe, bei der Einheiten systematisch ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zum Umfang. Die Auswahl erfolgt ohne Zurücklegen.
- **PPS sequenziell.** Dies ist eine Methode für die erste Stufe, bei der Einheiten sequenziell ohne Zurücklegen ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zur Klumpengröße.
- **PPS Brewer.** Dies ist eine Methode für die erste Stufe, bei der aus jeder Schicht zwei Klumpen ohne Zurücklegen ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zur Klumpengröße. Damit diese Methode verwendet werden kann, muss eine Klumpenvariable angegeben werden.
- **PPS Murthy.** Dies ist eine Methode für die erste Stufe, bei der aus jeder Schicht zwei Klumpen ohne Zurücklegen ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zur Klumpengröße. Damit diese Methode verwendet werden kann, muss eine Klumpenvariable angegeben werden.
- **PPS Sampford.** Dies ist eine Methode für die erste Stufe, bei der aus jeder Schicht mehr als zwei Klumpen ohne Zurücklegen ausgewählt werden; die Auswahlwahrscheinlichkeit ist proportional zur Klumpengröße. Es handelt sich um eine Erweiterung der Brewer-Methode. Damit diese Methode verwendet werden kann, muss eine Klumpenvariable angegeben werden.
- **MZ-Schätzung für Analyse verwenden.** Standardmäßig wird in der Plandatei eine Schätzmethode angegeben, die mit der ausgewählten Stichprobenmethode konsistent ist. Dadurch können Sie eine Schätzung mit Zurücklegen verwenden, selbst wenn die Stichprobenmethode eine Schätzung ohne Zurücklegen beinhaltet. Diese Option ist nur in Stufe 1 verfügbar.

Maß für die Größe der Einheiten (MOS). Bei Auswahl einer PPS-Methode müssen Sie ein Maß für die Größe angeben, mit dem die Größe jeder Einheit festgelegt wird. Diese Größen können explizit in einer Variablen definiert oder aus den Daten berechnet werden. Optional können Sie für das MOS Unter- und Obergrenzen festlegen, die Vorrang vor allen Werten haben, die in der MOS-Variablen gefunden oder aus den Daten berechnet werden. Diese Optionen sind nur in Stufe 1 verfügbar.

Stichprobenassistent: Stichprobenumfang

Abbildung 2-4
Stichprobenassistent – Schritt "Stichprobenumfang"



In diesem Schritt können Sie die Anzahl bzw. den Anteil der Einheiten festlegen, die in der aktuellen Stufe als Stichprobe gezogen werden sollen. Der Stichprobenumfang kann fest oder für die verschiedenen Schichten unterschiedlich sein. Für die Angabe des Stichprobenumfangs können die in den vorangegangenen Stufen ausgewählten Klumpen verwendet werden, um Schichten zu definieren.

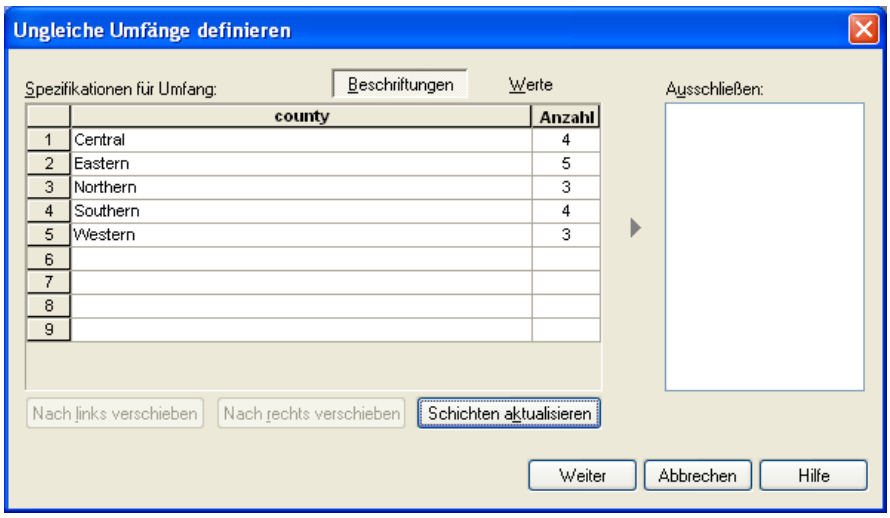
Einheiten. Sie können einen genauen Stichprobenumfang oder den Anteil der Einheiten für die Stichprobe angeben.

- **Wert.** Allen Schichten wird derselbe Wert zugewiesen. Wenn Anzahl als Metrik für die Einheiten ausgewählt wurde, sollten Sie eine positive Ganzzahl eingeben. Bei Auswahl von Anteile sollten Sie einen nichtnegativen Wert eingeben. Außer bei Stichprobenziehung mit Zurücklegen dürfen die Anteilswerte außerdem nicht größer als 1 sein.
- **Ungleiche Werte für Schichten.** Ermöglicht die Eingabe von Umfangswerten für die einzelnen Schichten über das Dialogfeld "Ungleiche Werte für Schichten".
- **Werte aus Variable lesen.** Ermöglicht die Auswahl einer numerischen Variablen, die die Umfangswerte für Schichten enthält.

Bei Auswahl von Anteile haben Sie die Möglichkeit, Unter- und Obergrenzen für die Anzahl der in der Stichprobe enthaltenen Einheiten festzulegen.

Ungleiche Umfänge definieren

Abbildung 2-5
Dialogfeld "Ungleiche Umfänge definieren"



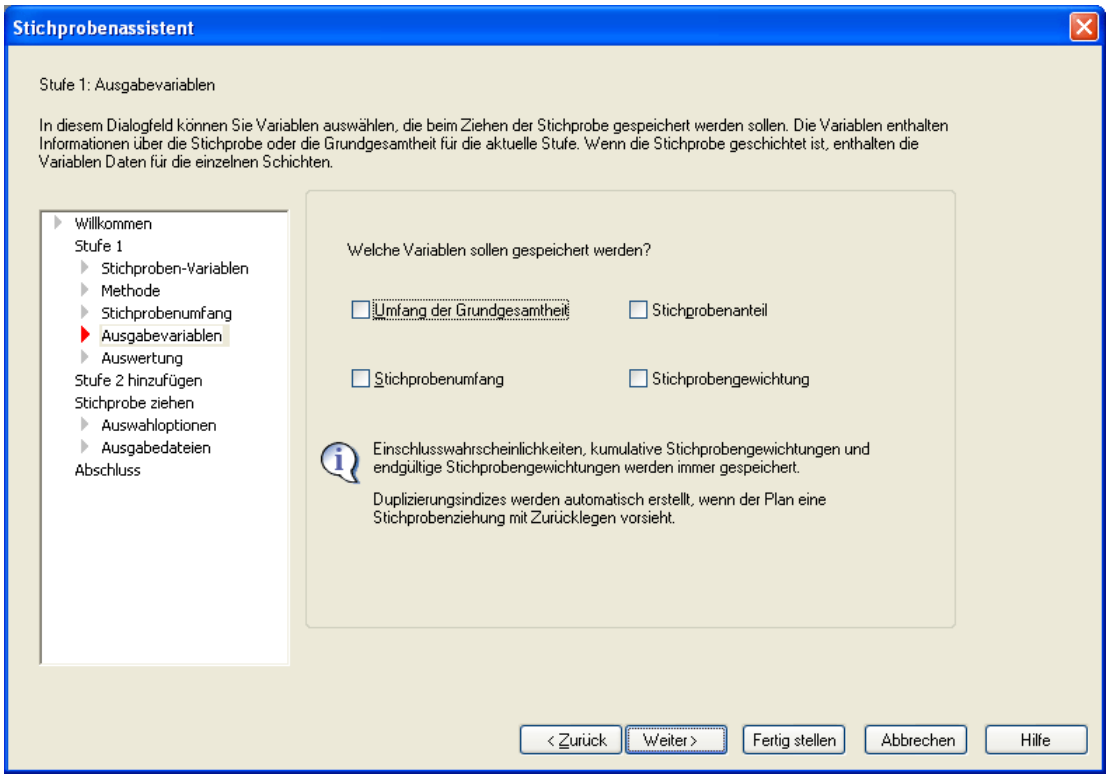
Im Dialogfeld "Ungleiche Umfänge definieren" können Sie Umfänge schichtweise eingeben.

Gitter "Spezifikationen für Umfang". Das Gitter zeigt die Kombination von bis zu fünf Schicht- oder Klumpenvariablen an, in jeder Zeile eine Schicht-/Klumpenkombination. Als Gittervariablen zulässig sind alle Schichtungsvariablen aus der aktuellen Stufe und den vorangegangenen Stufen sowie alle Klumpenvariablen aus den vorangegangenen Stufen. Die Variablen können im Gitter neu angeordnet oder in die Ausschlussliste verschoben werden. Geben Sie die Werte für den Umfang in die Spalte rechts außen ein. Klicken Sie auf Labels (Beschriftungen) oder Werte, um die Anzeige der Wertelabels und der Datenwerte für die Schichtungs- und Klumpenvariablen in den Gitterzellen ein- bzw. auszuschalten. Bei Zellen, die Werte ohne Labels enthalten, werden immer Werte angezeigt. Klicken Sie auf Schichten aktualisieren, um das Gitter mit allen Kombinationen von beschrifteten Datenwerten für Variablen im Gitter neu auszufüllen.

Ausschließen. Um die Umfänge für eine Teilmenge von Schicht-/Klumpenkombinationen anzugeben, verschieben Sie eine oder mehrere Variablen in die Ausschlussliste. Diese Variablen werden nicht für die Festlegung der Stichprobenumfänge verwendet.

Stichprobenassistent: Ausgabevariablen

Abbildung 2-6
Stichprobenassistent – Schritt "Ausgabevariablen"



In diesem Schritt können Sie Variablen auswählen, die beim Ziehen der Stichprobe gespeichert werden sollen.

Umfang der Grundgesamtheit. Die geschätzte Anzahl an Einheiten in der Grundgesamtheit für eine bestimmte Stufe. Der Stammname der gespeicherten Variablen lautet *PopulationSize_*.

Stichprobenanteil. Die Stichprobenquote in einer bestimmten Stufe. Der Stammname der gespeicherten Variablen lautet *SamplingRate_*.

Stichprobenumfang. Die Anzahl der Einheiten, die in einer bestimmten Stufe als Stichprobe gezogen wurden. Der Stammname der gespeicherten Variablen lautet *SampleSize_*.

Stichprobengewichtung. Dies ist die Inverse der Einschlusswahrscheinlichkeiten. Der Stammname der gespeicherten Variablen lautet *SampleWeight_*.

Einige stufenweise Variablen werden automatisch generiert. Dazu gehören:

Einschlusswahrscheinlichkeiten. Der Anteil der Einheiten, die in einer bestimmten Stufe als Stichprobe gezogen wurden. Der Stammname der gespeicherten Variablen lautet *InclusionProbability_*.

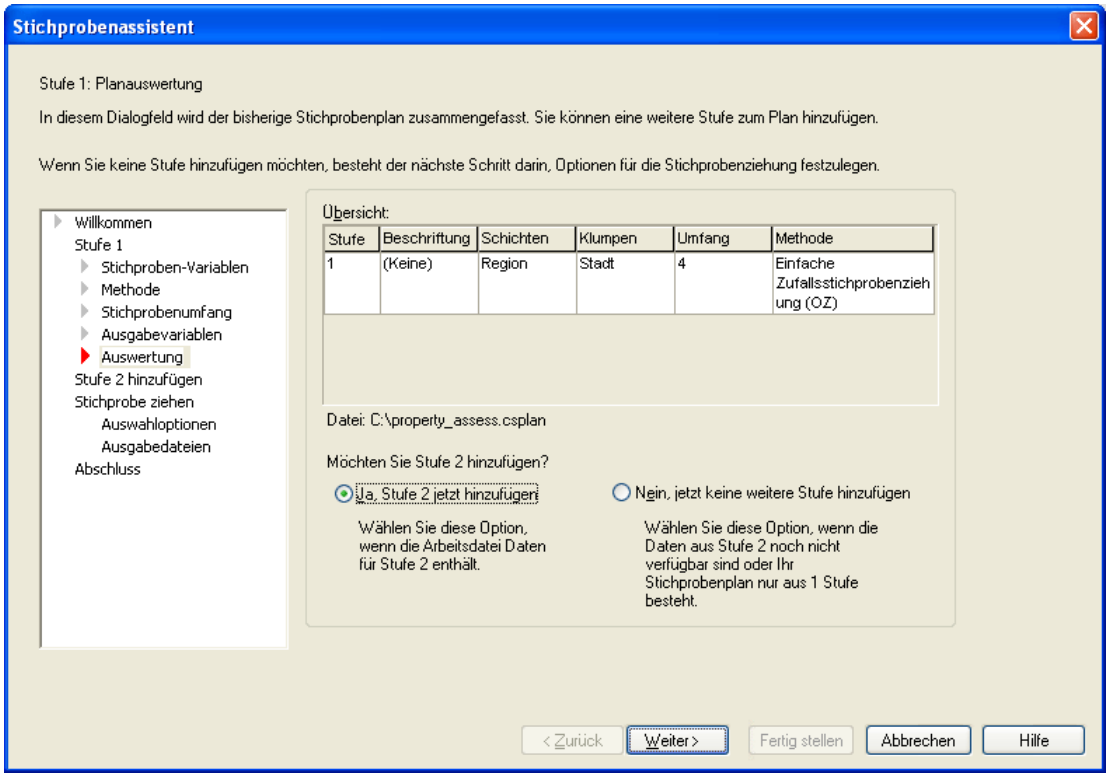
Kumulierte Gewichtung. Die kumulierte Stichprobengewichtung über alle Stufen einschließlich der aktuellen. Der Stammname der gespeicherten Variablen lautet *SampleWeightCumulative_*.

Index. Identifiziert Einheiten, die mehrmals in einer Stufe ausgewählt wurden. Der Stammname der gespeicherten Variablen lautet *Index_*.

Anmerkung: Die Stammnamen der gespeicherten Variablen beinhalten ein ganzzahliges Suffix, das der Stufennummer entspricht, beispielsweise *PopulationSize_1_* für die gespeicherte Größe des Stichprobenumfangs für Stufe 1.

Stichprobenassistent: Planübersicht

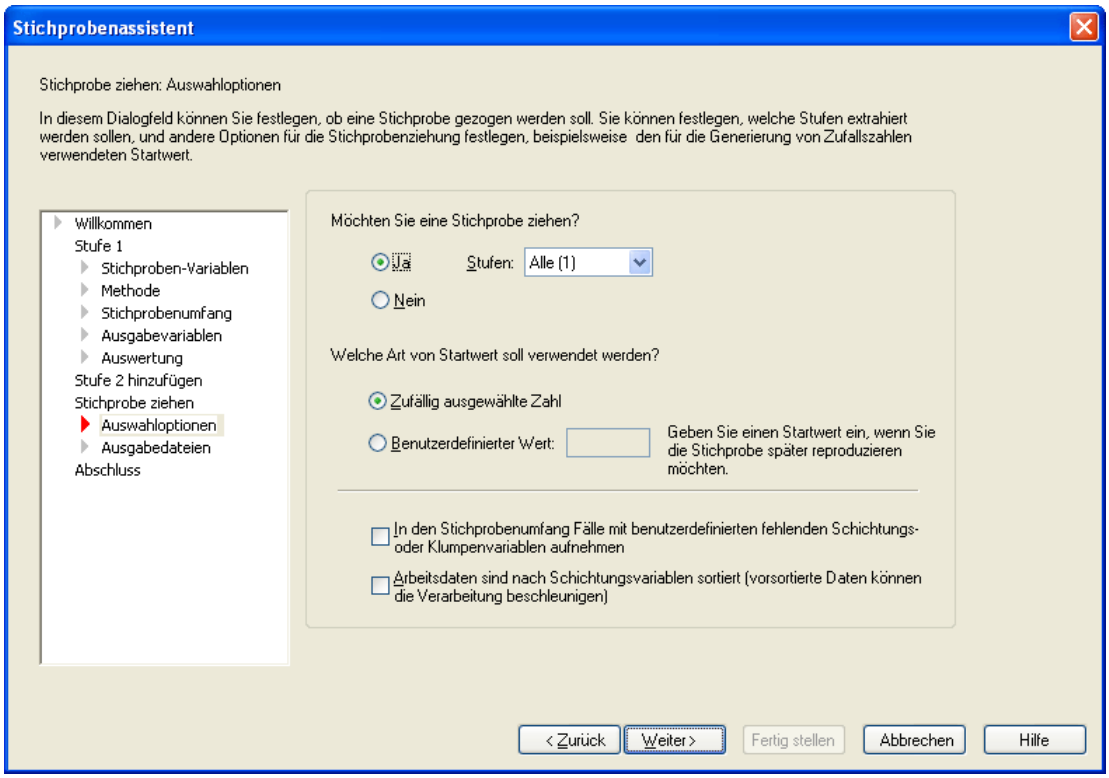
Abbildung 2-7
Stichprobenassistent – Schritt "Planübersicht"



Dies ist der letzte Schritt in jeder Stufe, in der eine Übersicht der in der aktuellen Stufe vorgenommenen Angaben zum Stichprobenplan angezeigt wird. Anschließend können Sie entweder zur nächsten Stufe weitergehen (und sie, falls erforderlich, erstellen) oder die Optionen für die Stichprobenziehung festlegen.

Stichprobenassistent: Stichprobe ziehen: Auswahloptionen

Abbildung 2-8
Stichprobenassistent – Schritt "Stichprobe ziehen: Auswahloptionen"


Stichprobenassistent
Stichprobe ziehen: Auswahloptionen
In diesem Dialogfeld können Sie festlegen, ob eine Stichprobe gezogen werden soll. Sie können festlegen, welche Stufen extrahiert werden sollen, und andere Optionen für die Stichprobenziehung festlegen, beispielsweise den für die Generierung von Zufallszahlen verwendeten Startwert.
Willkommen
Stufe 1
Stichproben-Variablen
Methode
Stichprobenumfang
Ausgabevariablen
Auswertung
Stufe 2 hinzufügen
Stichprobe ziehen
Auswahloptionen
Ausgabedateien
Abschluss
Möchten Sie eine Stichprobe ziehen?
Ja
Stufen:
Alle (1)
Nein
Welche Art von Startwert soll verwendet werden?
Zufällig ausgewählte Zahl
Benutzerdefinierter Wert:
Geben Sie einen Startwert ein, wenn Sie die Stichprobe später reproduzieren möchten.
In den Stichprobenumfang Fälle mit benutzerdefinierten fehlenden Schichtungs- oder Klumpenvariablen aufnehmen
Arbeitsdaten sind nach Schichtungsvariablen sortiert (vorsortierte Daten können die Verarbeitung beschleunigen)
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

In diesem Schritt können Sie auswählen, ob eine Stichprobe gezogen werden soll. Außerdem können Sie andere Optionen für die Stichprobenauswahl festlegen, beispielsweise den zufälligen Startwert und den Umgang mit fehlenden Werten.

Stichprobe ziehen. Sie können auswählen, ob eine Stichprobe gezogen werden soll. Außerdem können Sie festlegen, dass ein Teil des Stichprobenplans ausgeführt werden soll. Die Stichprobenziehungen für die Stufen muss der Reihe nach erfolgen. Die Stichprobenziehung für Stufe 2 kann also erst erfolgen, wenn die Stichprobe für Stufe 1 gezogen wurde. Bei der Bearbeitung oder Ausführung eines Plans kann für gesperrte Stufen keine erneute Stichprobenziehung erfolgen.

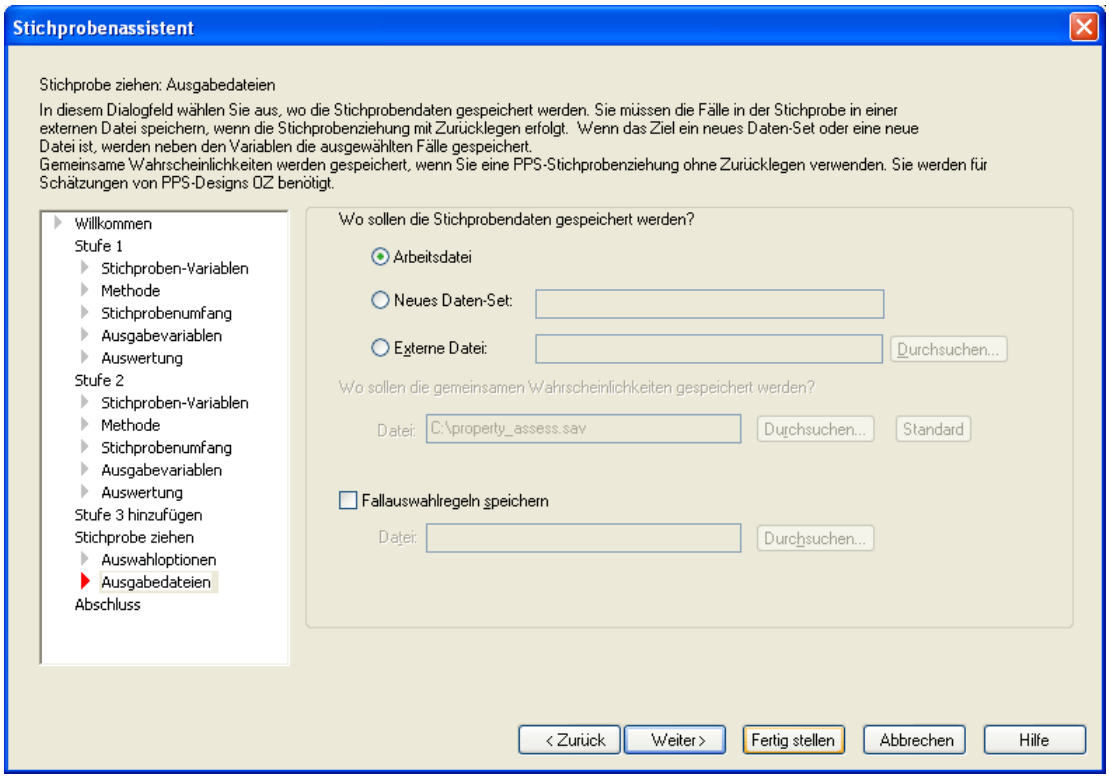
Startwert. Damit können Sie einen Startwert für die Generierung von Zufallszahlen auswählen.

Benutzerdefinierte fehlende Werte einschließen. Hiermit wird festgelegt, ob benutzerdefinierte fehlende Werte gültig sind. Wenn ja, werden benutzerdefinierte fehlende Werte als gesonderte Kategorie behandelt.

Daten sind bereits sortiert. Wenn der Stichprobenrahmen anhand der Werte der Schichtungsvariablen vorsortiert wurde, können Sie mit dieser Option den Auswahlvorgang beschleunigen.

Stichprobenassistent: Stichprobe ziehen: Ausgabedateien

Abbildung 2-9
Stichprobenassistent – Schritt "Stichprobe ziehen: Ausgabedateien"



In diesem Schritt können Sie die Ausgabedateien für die in der Stichprobe enthaltenen Fälle, die Gewichtungsvariablen, gemeinsamen Wahrscheinlichkeiten und Fallauswahlregeln angeben.

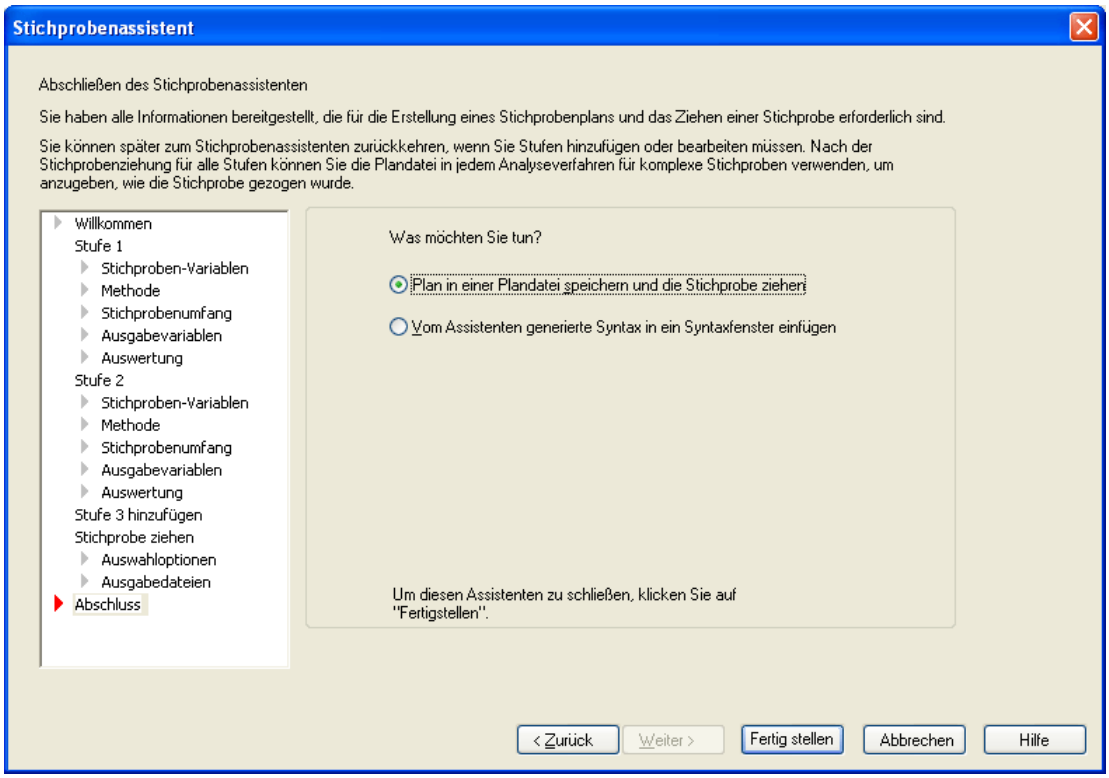
Stichprobendaten. Mit diesen Optionen können Sie festlegen, wo die Ausgabedaten für die Stichprobe gespeichert werden sollen. Sie können der Arbeitsdatei hinzugefügt, in ein neues Daten-Set geschrieben oder in einer externen Datendatei im SPSS-Format gespeichert werden. Daten-Sets sind in der aktuellen Sitzung verfügbar. In nachfolgenden Sitzungen stehen Daten-Sets nur zur Verfügung, wenn Sie sie ausdrücklich als Datendateien speichern. Die Namen von Daten-Sets müssen den Regeln zum Benennen von Variablen entsprechen. Wenn eine externe Datei oder ein neues Daten-Set angegeben wird, werden die Ausgabevariablen für die Stichprobenziehung und die Variablen in der Arbeitsdatei für die ausgewählten Fälle gespeichert.

Gemeinsame Wahrscheinlichkeiten. Mit diesen Optionen können Sie festlegen, in welche Datei die gemeinsamen Wahrscheinlichkeiten geschrieben werden sollen. Sie werden in einer externen Datendatei im SPSS-Format gespeichert. Gemeinsame Wahrscheinlichkeiten werden erstellt, wenn die PPS-Methode ohne Zurücklegen, "PPS Brewer", "PPS Sampford" oder "PPS Murthy" ausgewählt und "MZ-Schätzung" nicht angegeben ist.

Fallauswahlregeln. Wenn Sie die Stichprobe für jede Stufe einzeln konstruieren, sollten Sie die Fallauswahlregeln in einer Textdatei speichern. Die Fallauswahlregeln unterstützen die Erstellung des Teilrahmens für die weiteren Stufen.

Stichprobenassistent: Fertig stellen

Abbildung 2-10
Stichprobenassistent – Schritt "Fertig stellen"


Stichprobenassistent
Abschließen des Stichprobenassistenten
Sie haben alle Informationen bereitgestellt, die für die Erstellung eines Stichprobenplans und das Ziehen einer Stichprobe erforderlich sind.
Sie können später zum Stichprobenassistenten zurückkehren, wenn Sie Stufen hinzufügen oder bearbeiten müssen. Nach der Stichprobenziehung für alle Stufen können Sie die Plandatei in jedem Analyseverfahren für komplexe Stichproben verwenden, um anzugeben, wie die Stichprobe gezogen wurde.
Willkommen
Stufe 1
Stichproben-Variablen
Methode
Stichprobenumfang
Ausgabevariablen
Auswertung
Stufe 2
Stichproben-Variablen
Methode
Stichprobenumfang
Ausgabevariablen
Auswertung
Stufe 3 hinzufügen
Stichprobe ziehen
Auswahloptionen
Ausgabedateien
Abschluss
Was möchten Sie tun?
Plan in einer Plandatei speichern und die Stichprobe ziehen
Vom Assistenten generierte Syntax in ein Syntaxfenster einfügen
Um diesen Assistenten zu schließen, klicken Sie auf "Fertigstellen".
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Dies ist der letzte Schritt. Sie können die Plandatei jetzt speichern und die Stichprobe ziehen oder Ihre Auswahl in ein Syntax-Fenster einfügen.

Wenn Sie Änderungen an Schritten in der bestehenden Plandatei vornehmen, können Sie den bearbeiteten Plan in einer neuen Datei speichern oder die bestehende Plandatei überschreiben. Wenn Sie Stufen hinzufügen, ohne Änderungen an bestehenden Stufen vorzunehmen, überschreibt der Assistent die bestehende Plandatei. Wenn Sie den Plan in einer neuen Datei speichern möchten, wählen Sie die Option Vom Assistenten generierte Syntax in ein Syntaxfenster einfügen aus und ändern Sie den Dateinamen in den Syntaxbefehlen.

Bearbeiten eines bestehenden Stichprobenplans

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Komplexe Stichproben
  Stichprobe auswählen...

- Wählen Sie die Option Stichprobenplan bearbeiten und wählen Sie eine Plandatei zur Bearbeitung aus.

- Klicken Sie auf Weiter, um unter Verwendung des Assistenten fortzufahren.

▶ Überarbeiten Sie den Stichprobenplan im Schritt “Planübersicht” und klicken Sie auf Weiter.

Die darauf folgenden Schritte sind größtenteils mit denen für einen neuen Plan identisch. Weitere Informationen finden Sie in der Hilfe zu den einzelnen Schritten.

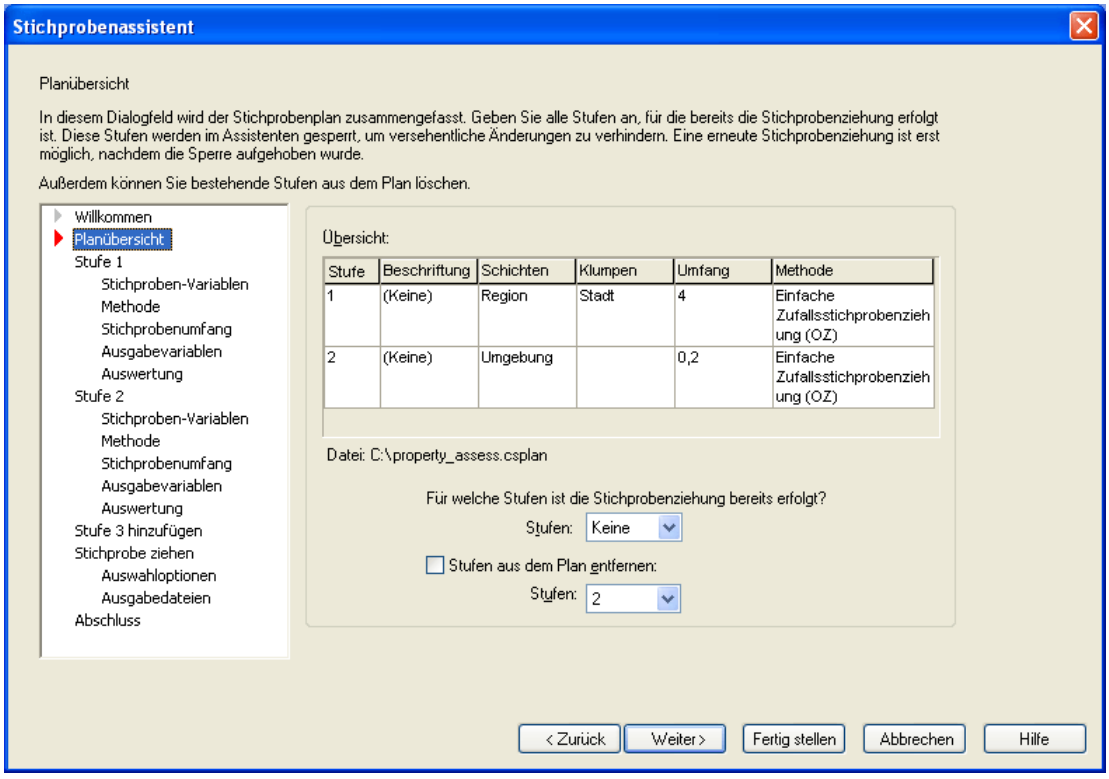
▶ Wechseln Sie zum Schritt “Fertig stellen” und geben Sie einen neuen Namen für die bearbeitete Plandatei an oder legen Sie fest, dass die bestehende Plandatei überschrieben werden soll.

Die folgenden Optionen sind verfügbar:

- Angabe der Stufen, für die die Stichprobenziehung bereits erfolgt ist.
- Stufen aus dem Plan entfernen.

Stichprobenassistent: Planübersicht

Abbildung 2-11
Stichprobenassistent – Schritt “Planübersicht”



In diesem Schritt können Sie den Stichprobenplan überprüfen und Stufen angeben, für die die Stichprobenziehung bereits erfolgt ist. Wenn Sie einen Plan bearbeiten, können Sie auch Stufen aus dem Plan entfernen.

Stufen mit erfolgter Stichprobenziehung. Wenn kein erweiterter Stichprobenrahmen verfügbar ist, müssen Sie einen mehrstufigen Stichprobenplan für jede Stufe einzeln ausführen. Wählen Sie in der Dropdown-Liste aus, für welche Stufen die Stichprobenziehung bereits erfolgt ist. Alle Stufen, die bereits ausgeführt wurden, sind gesperrt, stehen also im Schritt “Stichprobe ziehen:

Auswahloptionen” nicht zur Verfügung und können beim Bearbeiten des Plans nicht geändert werden.

Stufen entfernen. Sie können die Stufen 2 und 3 aus einem mehrstufigen Plan entfernen.

Ausführen eines bestehenden Stichprobenplans

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Komplexe Stichproben
  Stichprobe auswählen...

- Wählen Sie die Option Stichprobe ziehen und wählen Sie eine Plandatei für die Ausführung aus.
- Klicken Sie auf Weiter, um unter Verwendung des Assistenten fortzufahren.
- Überarbeiten Sie den Stichprobenplan im Schritt “Planübersicht” und klicken Sie auf Weiter.
- Die einzelnen Schritte mit Informationen zur Stufe werden bei der Ausführung eines Stichprobenplans übersprungen. Sie können nun jederzeit mit dem Schritt “Fertig stellen” fortfahren.

  Optional können Sie die Stufen angeben, für die die Stichprobenziehung bereits erfolgt ist.

Zusätzliche Funktionen bei den Befehlen CSPLAN und CSSELECT

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

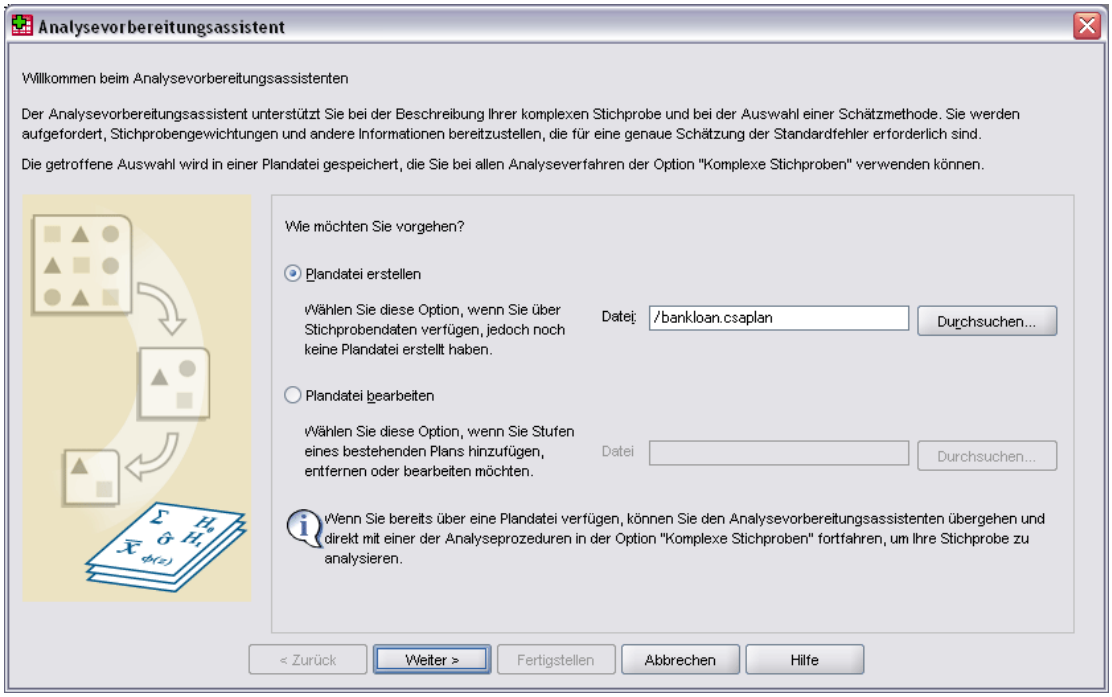
- Festlegen der benutzerdefinierten Namen für Ausgabevariablen.
- Festlegen der Ausgabe im Viewer. Sie können beispielsweise die stufenweise Übersicht über den Plan unterdrücken, der angezeigt wird, wenn eine Stichprobe entworfen oder bearbeitet wird, die Übersicht über die Fälle in der Stichprobe für die einzelnen Schichten unterdrücken und eine Zusammenfassung der Fallverarbeitung abrufen.
- Wählen Sie eine Teilmenge der Variablen in der Arbeitsdatei aus, die in eine externe Stichprobendatei oder in ein anderes Daten-Set geschrieben werden soll.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
3

Vorbereiten einer komplexen Stichprobe für die Analyse

Abbildung 3-1
Analysevorbereitungsassistent – Schritt "Willkommen"


Analysevorbereitungsassistent
Willkommen beim Analysevorbereitungsassistenten
Der Analysevorbereitungsassistent unterstützt Sie bei der Beschreibung Ihrer komplexen Stichprobe und bei der Auswahl einer Schätzmethode. Sie werden aufgefordert, Stichprobengewichtungen und andere Informationen bereitzustellen, die für eine genaue Schätzung der Standardfehler erforderlich sind.
Die getroffene Auswahl wird in einer Plandatei gespeichert, die Sie bei allen Analyseverfahren der Option "Komplexe Stichproben" verwenden können.
Wie möchten Sie vorgehen?
Plandatei erstellen
Wählen Sie diese Option, wenn Sie über Stichprobendaten verfügen, jedoch noch keine Plandatei erstellt haben.
Datei:
/bankloan.csaplan
Durchsuchen...
Plandatei bearbeiten
Wählen Sie diese Option, wenn Sie Stufen eines bestehenden Plans hinzufügen, entfernen oder bearbeiten möchten.
Datei
Durchsuchen...
Wenn Sie bereits über eine Plandatei verfügen, können Sie den Analysevorbereitungsassistenten übergehen und direkt mit einer der Analyseprozeduren in der Option "Komplexe Stichproben" fortfahren, um Ihre Stichprobe zu analysieren.
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

Der Analysevorbereitungsassistent führt Sie durch die Schritte zum Erstellen bzw. Bearbeiten eines Analyseplans zur Verwendung mit den verschiedenen Analyseverfahren für komplexe Stichproben. Vor der Verwendung des Assistenten sollten Sie nach einem komplexen Plan eine Stichprobe gezogen haben.

Das Erstellen eines neuen Plans ist am sinnvollsten, wenn Sie keinen Zugriff auf die Datei mit dem Stichprobenplan haben, der zum Ziehen der Stichprobe verwendet wurde (der Stichprobenplan enthält einen Standard-Analyseplan). Wenn Sie Zugriff auf die Datei mit dem Stichprobenplan haben, der zum Ziehen der Stichprobe verwendet wurde, können Sie den in der Datei enthaltenen Standard-Analyseplan verwenden oder die Standardfestlegungen für die Analyse abändern und Ihre Änderungen in einer neuen Datei speichern.

Erstellen eines neuen Analyseplans

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Komplexe Stichproben
  Für Analyse vorbereiten...

- ▶ Wählen Sie die Option Plandatei erstellen aus und wählen Sie einen Dateinamen für die Plandatei, in der der Analyseplan gespeichert werden soll.

- ▶ Klicken Sie auf Weiter, um unter Verwendung des Assistenten fortzufahren.

- ▶ Geben Sie die Variable mit den Stichprobengewichten im Schritt “Stichproben-Variablen” an. Definieren Sie gegebenenfalls Schichten und Klumpen.

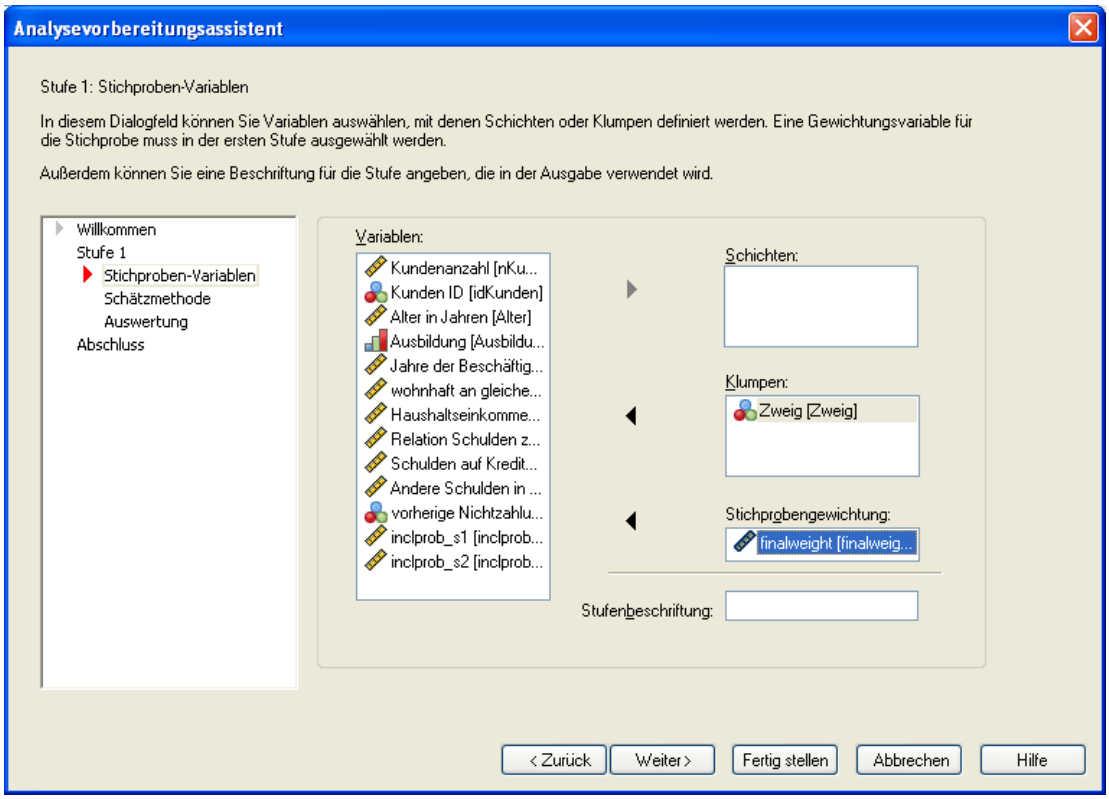
- ▶ Jetzt können Sie auf Fertig stellen klicken, um den Plan zu speichern.

  Optional können Sie in weiteren Schritten folgende Aktionen durchführen:

  - Auswahl der Methode zum Schätzen der Standardfehler im Schritt “Schätzmethode”.
  - Angabe der Anzahl der Einheiten in der Stichprobe oder der Einschlusswahrscheinlichkeit pro Einheit im Schritt “Umfang”.
  - Hinzufügen einer zweiten oder dritten Stufe zum Plan.
  - Einfügen der getroffenen Auswahl als Befehlssyntax.

Analysevorbereitungsassistent: Stichproben-Variablen

Abbildung 3-2
Analysevorbereitungsassistent – Schritt "Stichproben-Variablen"



In diesem Schritt können Sie die Schichtungs- und Klumpenvariablen identifizieren und Stichprobengewichte definieren. Außerdem können Sie eine Beschriftung für die Stufe angeben.

Schichten. Durch die Kombination von Schichtungsvariablen werden eindeutige Teilgesamtheiten, so genannte Schichten, definiert. Die gesamte Stichprobe besteht aus der Kombination unabhängiger Stichproben aus jeder Schicht.

Klumpen. Klumpenvariablen definieren Gruppen von Beobachtungseinheiten, so genannte Klumpen. Bei in mehreren Stufen gezogenen Stichproben werden Klumpen in den früheren Stufen und anschließend Teilstichproben aus den ausgewählten Klumpen gezogen. Beim Analysieren von Datendateien, die durch die Ziehung von Klumpenstichproben mit Zurücklegen erstellt werden, sollten Sie auch den Duplizierungsindex als Klumpenvariable verwenden.

Stichprobengewichtung. In der ersten Stufe müssen Sie Stichprobengewichte angeben. Die Stichprobengewichte für die weiteren Stufen des aktuellen Plans werden automatisch berechnet.

Stufenbeschriftung. Sie können für jede Stufe ein optionales String-Label angeben. Dieses wird in der Ausgabe verwendet, um die stufenweisen Informationen besser identifizieren zu können.

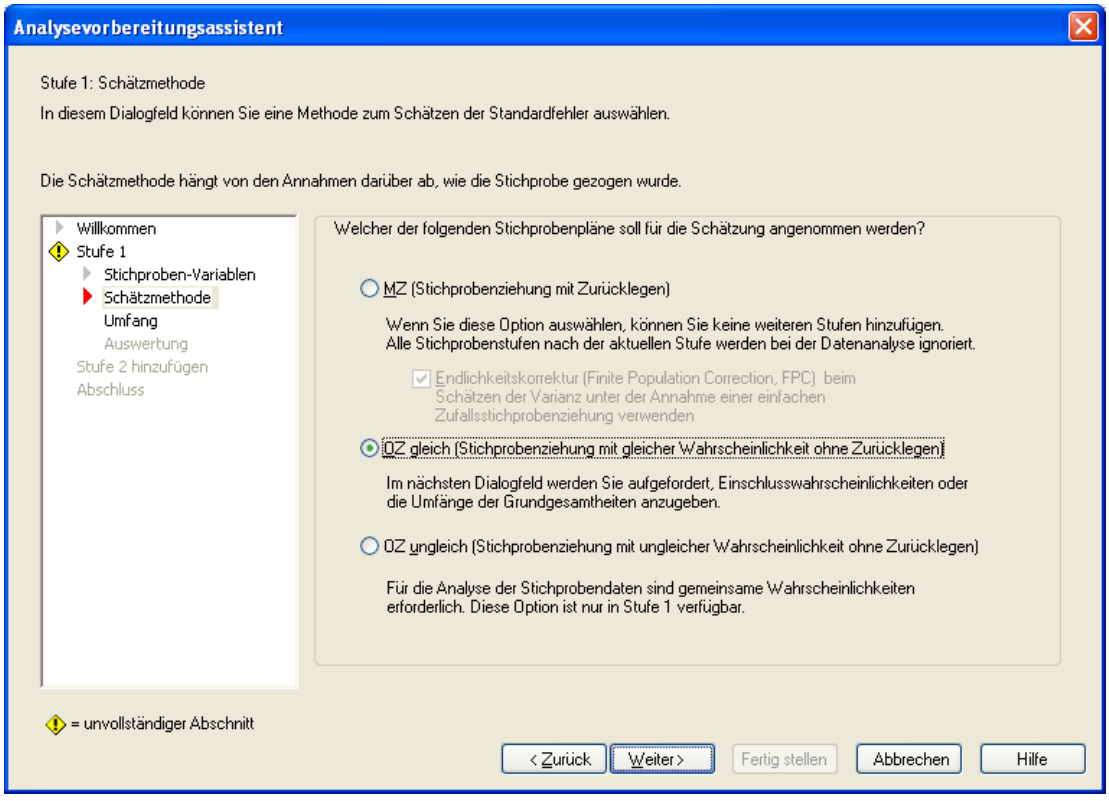
Anmerkung: Die Liste der Quellvariablen hat in allen Schritten des Assistenten denselben Inhalt. Anders ausgedrückt: Variablen, die in einem Schritt aus der Liste der Quellvariablen entfernt werden, werden in allen Schritten aus der Liste entfernt. Variablen, die wieder zur Liste der Quellvariablen hinzugefügt werden, erscheinen in allen Schritten.

Baumsteuerungen zur Navigation im Analyseassistenten

Auf der linken Seite jedes Schritts im Analyseassistenten finden Sie eine Gliederung, die eine Übersicht über alle Schritte bietet. Sie können im Assistenten navigieren, indem Sie in der Gliederung auf den Namen eines aktivierten Schrittes klicken. Schritte sind aktiviert, wenn alle vorangegangenen Schritte gültig sind – d. h. solange für jeden vorangegangen Schritt die erforderlichen Mindestangaben vorgenommen wurden. Weitere Informationen dazu, warum ein Schritt möglicherweise ungültig ist, finden Sie in der Hilfe zu den einzelnen Schritten.

Analysevorbereitungsassistent: Schätzmethode

Abbildung 3-3
Analysevorbereitungsassistent – Schritt "Schätzmethode"


Analysevorbereitungsassistent
Stufe 1: Schätzmethode
In diesem Dialogfeld können Sie eine Methode zum Schätzen der Standardfehler auswählen.
Die Schätzmethode hängt von den Annahmen darüber ab, wie die Stichprobe gezogen wurde.
Willkommen
Stufe 1
Stichproben-Variablen
Schätzmethode
Umfang
Auswertung
Stufe 2 hinzufügen
Abschluss
Welcher der folgenden Stichprobenpläne soll für die Schätzung angenommen werden?
MZ (Stichprobenziehung mit Zurücklegen)
Wenn Sie diese Option auswählen, können Sie keine weiteren Stufen hinzufügen.
Alle Stichprobenstufen nach der aktuellen Stufe werden bei der Datenanalyse ignoriert.
Endlichkeitskorrektur (Finite Population Correction, FPC) beim Schätzen der Varianz unter der Annahme einer einfachen Zufallsstichprobenziehung verwenden
OZ gleich (Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen)
Im nächsten Dialogfeld werden Sie aufgefordert, Einschlusswahrscheinlichkeiten oder die Umfänge der Grundgesamtheiten anzugeben.
OZ ungleich (Stichprobenziehung mit ungleicher Wahrscheinlichkeit ohne Zurücklegen)
Für die Analyse der Stichprobendaten sind gemeinsame Wahrscheinlichkeiten erforderlich. Diese Option ist nur in Stufe 1 verfügbar.
= unvollständiger Abschnitt
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

In diesem Schritt können Sie eine Schätzmethode für die Stufe angeben.

MZ (Stichprobenziehung mit Zurücklegen). Die MZ-Schätzung beinhaltet keine Endlichkeitskorrektur, (Finite Population Correction, FPC) bei der Schätzung der Varianz in einem komplexen Stichprobenplan. Beim Schätzen der Varianz bei einer einfachen

Zufallsstrichprobenziehung (Simple Random Sampling, SRS) können Sie auswählen, ob die Endlichkeitskorrektur (FPC) aufgenommen oder ausgeschlossen werden soll.

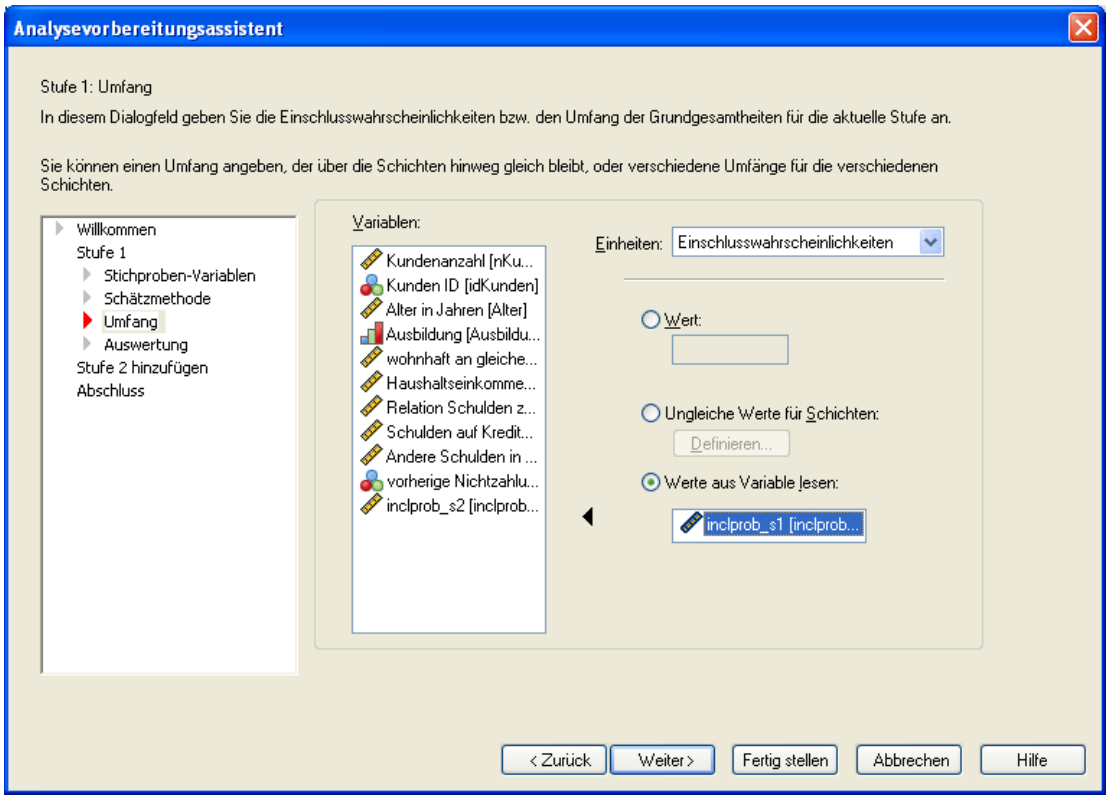
Es wird empfohlen, bei der SRS-Varianzschätzung keine FPC aufzunehmen, wenn die Analysegewichtungen skaliert wurden, sodass ihre Summe nicht die Populationsgröße ergibt. Der SRS-Varianzschätzer wird bei der Berechnung von Statistiken wie dem Effekt des Stichprobenplans verwendet. MZ-Schätzung kann nur in der letzten Stufe eines Planes angegeben werden; der Assistent lässt nicht zu, dass eine weitere Stufe hinzugefügt wird, wenn die Option für die MZ-Schätzung ausgewählt wird.

OZ gleich (Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen). Die Schätzung für "OZ gleich" beinhaltet eine Endlichkeitskorrektur und geht davon aus, dass die Stichprobenziehung bei den Einheiten mit gleicher Wahrscheinlichkeit erfolgt. "OZ gleich" kann in jeder Stufe eines Plans angegeben werden.

OZ ungleich (Stichprobenziehung mit ungleicher Wahrscheinlichkeit ohne Zurücklegen). Neben der Endlichkeitskorrektur berücksichtigt "OZ ungleich" auch Stichprobeneinheiten (in der Regel Klumpen), die mit ungleicher Wahrscheinlichkeit ausgewählt wurden. Diese Schätzmethode ist nur in der ersten Stufe verfügbar.

Analysevorbereitungsassistent: Umfang

Abbildung 3-4
Analysevorbereitungsassistent – Schritt "Umfang"



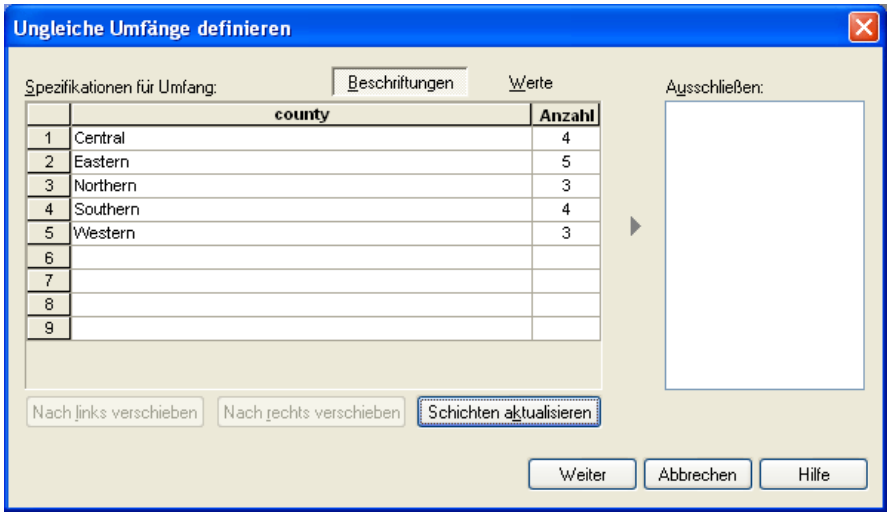
Dieser Schritt dient zur Angabe der Einschlusswahrscheinlichkeiten bzw. der Umfänge der Grundgesamtheiten für die aktuelle Stufe. Die Umfänge können fest oder für die verschiedenen Schichten unterschiedlich sein. Für die Angabe der Umfänge können die in den vorangegangenen Stufen festgelegten Klumpen verwendet werden, um Schichten zu definieren. Beachten Sie, dass dieser Schritt nur dann erforderlich ist, wenn als Schätzmethode “OZ gleich” ausgewählt wurde.

Einheiten. Sie können den genauen Umfang der Grundgesamtheiten angeben oder die Wahrscheinlichkeiten, mit denen die Stichprobenziehung der Einheiten erfolgte.

- **Wert.** Allen Schichten wird derselbe Wert zugewiesen. Wenn Umfang der Grundgesamtheiten als Metrik für die Einheiten ausgewählt wurde, sollten Sie eine nichtnegative Ganzzahl eingeben. Bei Auswahl von Einschlusswahrscheinlichkeiten sollten Sie einen Wert aus dem Bereich von 0 bis 1 eingeben.
- **Ungleiche Werte für Schichten.** Ermöglicht die Eingabe von Umfangswerten für die einzelnen Schichten über das Dialogfeld “Ungleiche Werte für Schichten”.
- **Werte aus Variable lesen.** Ermöglicht die Auswahl einer numerischen Variablen, die die Umfangswerte für Schichten enthält.

Ungleiche Umfänge definieren

Abbildung 3-5
Dialogfeld “Ungleiche Umfänge definieren”



Im Dialogfeld “Ungleiche Umfänge definieren” können Sie Umfänge schichtweise eingeben.

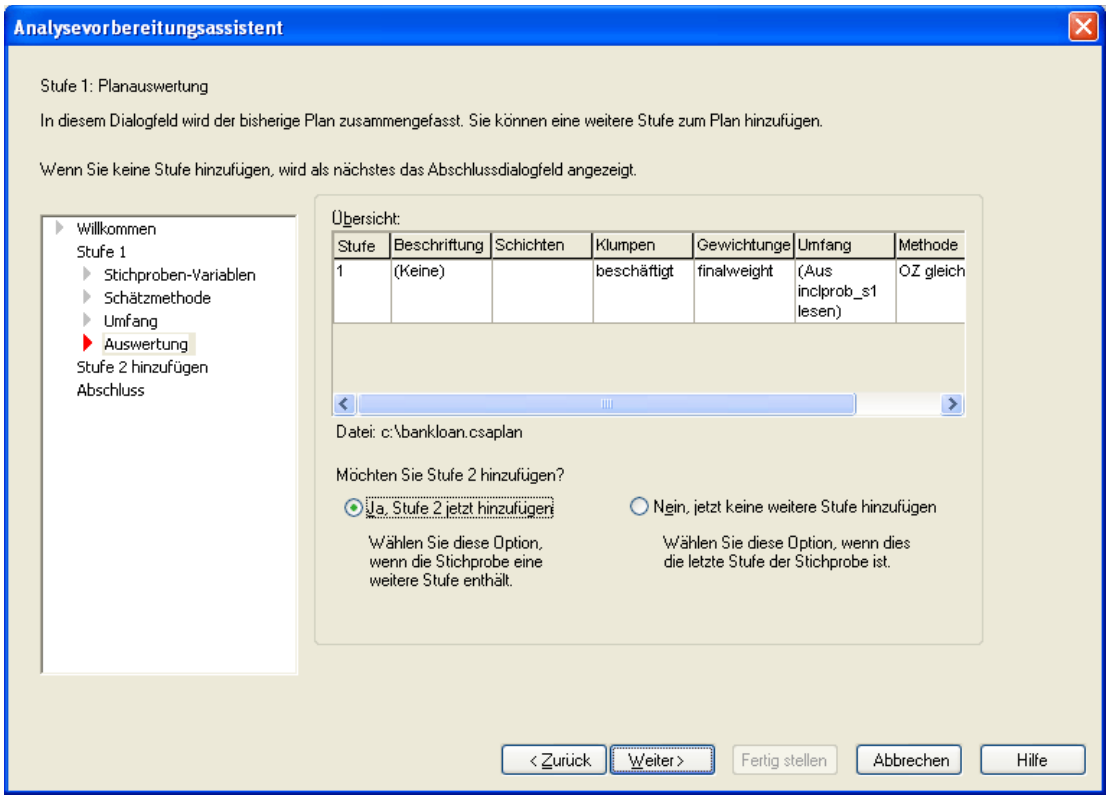
Gitter “Spezifikationen für Umfang”. Das Gitter zeigt die Kombination von bis zu fünf Schicht- oder Klumpenvariablen an, in jeder Zeile eine Schicht-/Klumpenkombination. Als Gittervariablen zulässig sind alle Schichtungsvariablen aus der aktuellen Stufe und den vorangegangenen Stufen sowie alle Klumpenvariablen aus den vorangegangenen Stufen. Die Variablen können im Gitter neu angeordnet oder in die Ausschlussliste verschoben werden. Geben Sie die Werte für den Umfang in die Spalte rechts außen ein. Klicken Sie auf Labels (Beschriftungen) oder Werte, um die Anzeige der Wertelabels und der Datenwerte für die Schichtungs- und Klumpenvariablen in den Gitterzellen ein- bzw. auszuschalten. Bei Zellen, die Werte ohne Labels enthalten, werden immer

Werte angezeigt. Klicken Sie auf Schichten aktualisieren, um das Gitter mit allen Kombinationen von beschrifteten Datenwerten für Variablen im Gitter neu auszufüllen.

Ausschließen. Um die Umfänge für eine Teilmenge von Schicht-/Klumpenkombinationen anzugeben, verschieben Sie eine oder mehrere Variablen in die Ausschlussliste. Diese Variablen werden nicht für die Festlegung der Stichprobenumfänge verwendet.

Analysevorbereitungsassistent: Planübersicht

Abbildung 3-6
Analysevorbereitungsassistent, Schritt "Planübersicht"



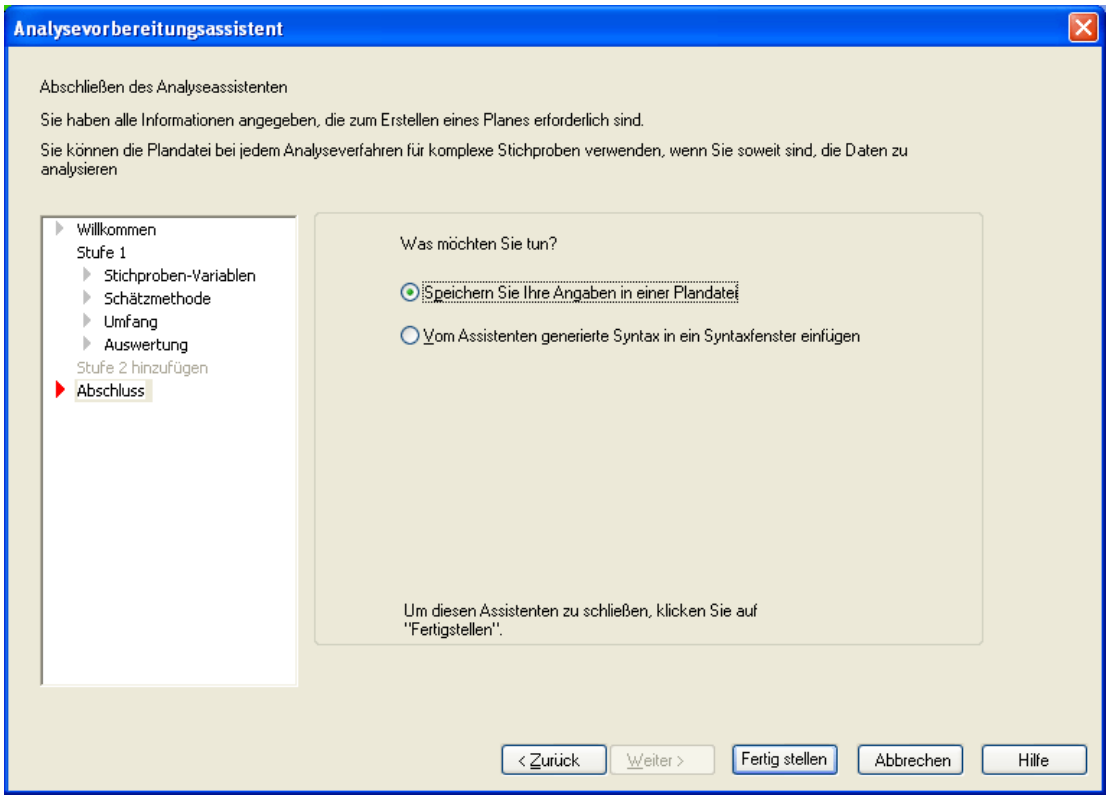
Dies ist der letzte Schritt in jeder Stufe, in dem eine Übersicht der in der aktuellen Stufe vorgenommenen Angaben zum Analyseplan angezeigt wird. Anschließend können Sie entweder zur nächsten Stufe weitergehen (und sie, falls erforderlich, erstellen) oder die Analyseangaben speichern.

Wenn keine weitere Stufe hinzugefügt werden kann, hat dies vermutlich einen der folgenden Gründe:

- Im Schritt "Stichproben-Variablen" wurde keine Klumpenvariable angegeben.
- Im Schritt "Schätzmethode" wurde "MZ-Schätzung" angegeben.
- Dies ist die dritte Stufe der Analyse, und der Assistent unterstützt maximal drei Stufen.

Analysevorbereitungsassistent: Fertig stellen

Abbildung 3-7
Analysevorbereitungsassistent – Schritt "Fertig stellen"


Analysevorbereitungsassistent
Abschließen des Analyseassistenten
Sie haben alle Informationen angegeben, die zum Erstellen eines Planes erforderlich sind.
Sie können die Plandatei bei jedem Analyseverfahren für komplexe Stichproben verwenden, wenn Sie soweit sind, die Daten zu analysieren
Willkommen
Stufe 1
Stichproben-Variablen
Schätzmethode
Umfang
Auswertung
Stufe 2 hinzufügen
Abschluss
Was möchten Sie tun?
Speichern Sie Ihre Angaben in einer Plandatei
Vom Assistenten generierte Syntax in ein Syntaxfenster einfügen
Um diesen Assistenten zu schließen, klicken Sie auf "Fertigstellen".
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Dies ist der letzte Schritt. Sie können die Plandatei jetzt speichern oder Ihre Auswahl in ein Syntax-Fenster einfügen.

Wenn Sie Änderungen an Schritten in der bestehenden Plandatei vornehmen, können Sie den bearbeiteten Plan in einer neuen Datei speichern oder die bestehende Plandatei überschreiben. Wenn Sie Stufen hinzufügen, ohne Änderungen an bestehenden Stufen vorzunehmen, überschreibt der Assistent die bestehende Plandatei. Wenn Sie den Plan in einer neuen Datei speichern möchten, wählen Sie die Option Vom Assistenten generierte Syntax in ein Syntaxfenster einfügen aus und ändern Sie den Dateinamen in den Syntaxbefehlen.

Bearbeiten eines bestehenden Analyseplans

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Komplexe Stichproben
  Für Analyse vorbereiten...

- Wählen Sie die Option Plandatei bearbeiten aus und wählen Sie einen Dateinamen für die Plandatei, in der der Analyseplan gespeichert werden soll.

- Klicken Sie auf Weiter, um unter Verwendung des Assistenten fortzufahren.

- Überarbeiten Sie den Analyseplan im Schritt “Planübersicht” und klicken Sie auf **Weiter**.

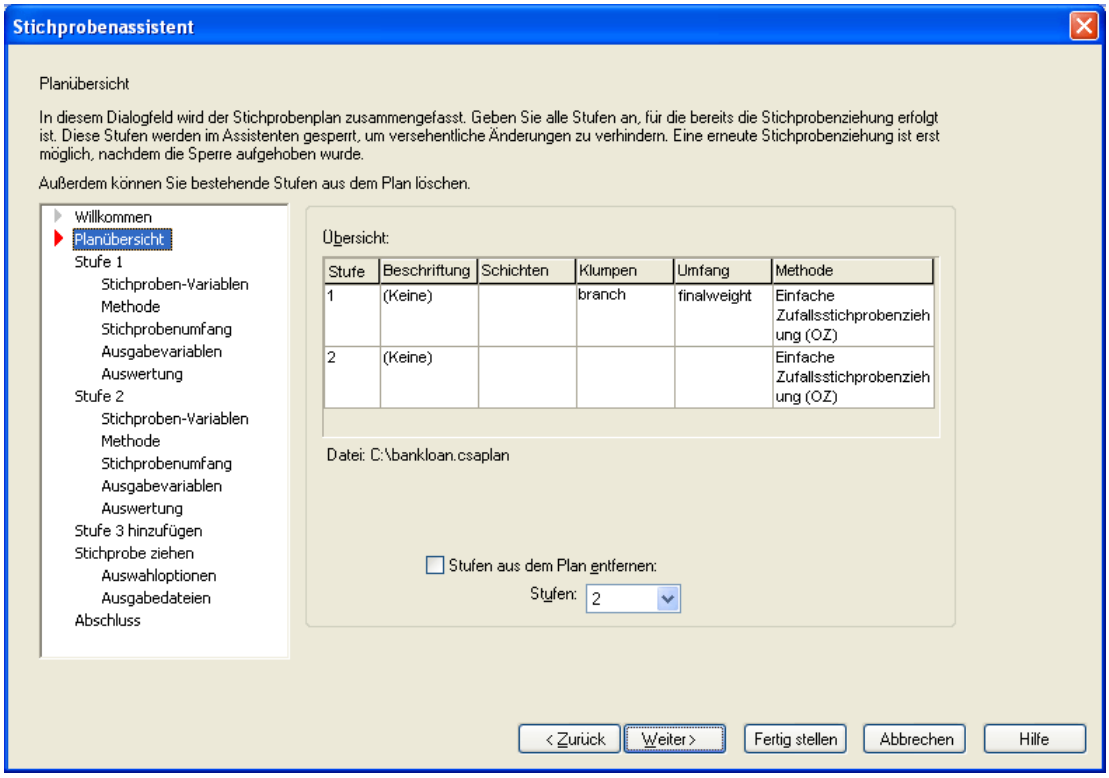
  Die darauf folgenden Schritte sind größtenteils mit denen für einen neuen Plan identisch. Weitere Informationen finden Sie in der Hilfe zu den einzelnen Schritten.

- Wechseln Sie zum Schritt “Fertig stellen” und geben Sie einen neuen Namen für die bearbeitete Plandatei an oder legen Sie fest, dass die bestehende Plandatei überschrieben werden soll.

  Optional können Sie Stufen aus dem Plan entfernen.

Analysevorbereitungsassistent: Planübersicht

Abbildung 3-8
Analysevorbereitungsassistent, Schritt “Planübersicht”



In diesem Schritt können Sie den Analyseplan überprüfen und Stufen aus dem Plan entfernen.

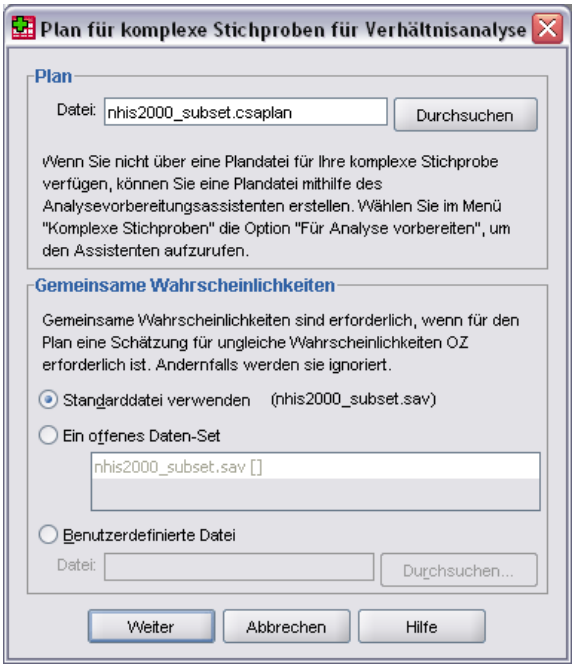
Stufen entfernen. Sie können die Stufen 2 und 3 aus einem mehrstufigen Plan entfernen. Da ein Plan mindestens eine Stufe aufweisen muss, können Sie die Stufe 1 zwar bearbeiten, nicht jedoch aus dem Plan entfernen.

Kapitel
4

Plan für komplexe Stichproben

Für die Analyseverfahren für komplexe Stichproben sind Analysespezifikationen aus einer Analyse- oder Stichprobenplan-Datei erforderlich, um gültige Ergebnisse zu erzielen.

Abbildung 4-1
Dialogfeld "Komplexe Stichproben: Plan"


Plan für komplexe Stichproben für Verhältnisanalyse
Plan
Datei:
nhis2000_subset.csaplan
Durchsuchen
Wenn Sie nicht über eine Plandatei für Ihre komplexe Stichprobe verfügen, können Sie eine Plandatei mithilfe des Analysevorbereitungsassistenten erstellen. Wählen Sie im Menü "Komplexe Stichproben" die Option "Für Analyse vorbereiten", um den Assistenten aufzurufen.
Gemeinsame Wahrscheinlichkeiten
Gemeinsame Wahrscheinlichkeiten sind erforderlich, wenn für den Plan eine Schätzung für ungleiche Wahrscheinlichkeiten OZ erforderlich ist. Andernfalls werden sie ignoriert.
Standarddatei verwenden (nhis2000_subset.sav)
Ein offenes Daten-Set
nhis2000_subset.sav []
Benutzerdefinierte Datei
Datei:
Durchsuchen...
Weiter
Abbrechen
Hilfe

Plan. Geben Sie den Pfad zu einer Analyse- oder Beispielplan-Datei an.

Gemeinsame Wahrscheinlichkeiten. Um die Schätzung "OZ ungleich" für Klumpen verwenden zu können, die mit einer PPS-Methode OZ gezogen wurden, müssen Sie eine gesonderte Datei oder ein geöffnetes Daten-Set angeben, die bzw. das die gemeinsamen Wahrscheinlichkeiten enthält. Diese Datei bzw. dieses Daten-Set wird vom Stichprobenassistenten während der Stichprobenziehung erstellt.

Kapitel
5

Häufigkeiten für komplexe Stichproben

Mit der Prozedur “Häufigkeiten für komplexe Stichproben” können Sie Häufigkeitstabellen für ausgewählte Variablen erstellen und univariate Statistiken anzeigen. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Beispiel. Mit der Prozedur “Häufigkeiten für komplexe Stichproben” können Sie univariate Statistiken in Tabellenform für die Einnahme von Vitaminpräparaten bei US-Bürgern erstellen, die auf den Ergebnissen der Umfrage National Health Interview Survey (NHIS) beruhen und einen geeigneten Analyseplan für diese öffentlich zugänglichen Daten beinhalten.

Statistiken. Mit diesem Verfahren erhalten Sie Schätzungen für die Umfänge der Grundgesamtheiten für die Zellen und Tabellenprozentsätze, außerdem Standardfehler, Konfidenzintervalle, Variationskoeffizienten, Effekte des Stichprobenplans, Quadratwurzeln aus den Effekten des Stichprobenplans, kumulative Werte sowie die ungewichtete Anzahl für jede Schätzung. Des Weiteren werden die Chi-Quadrat-Statistik und die Likelihood-Quotienten-Statistik für den Test auf gleiche Spaltenanteile berechnet.

Daten. Variablen, für die Häufigkeitstabellen erstellt werden, sollten kategorial sein. Bei den Variablen für die Teilgesamtheiten kann es sich um String-Variablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Plan, deren Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld “Komplexe Stichproben: Plan” ausgewählt wurde.

Berechnen von Häufigkeiten für komplexe Stichproben

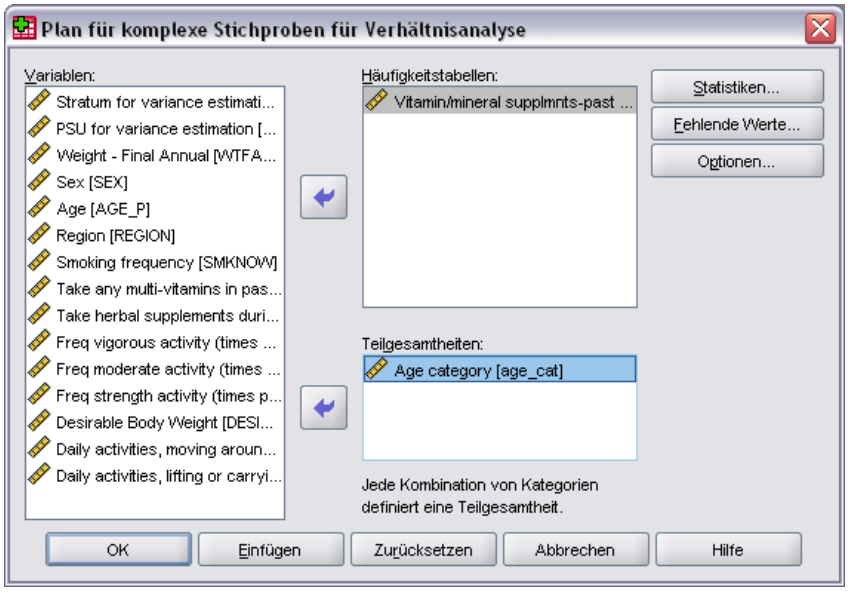
▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Komplexe Stichproben
    Häufigkeiten...

▶ Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.

▶ Klicken Sie auf Weiter.

Abbildung 5-1
Dialogfeld "Häufigkeiten"

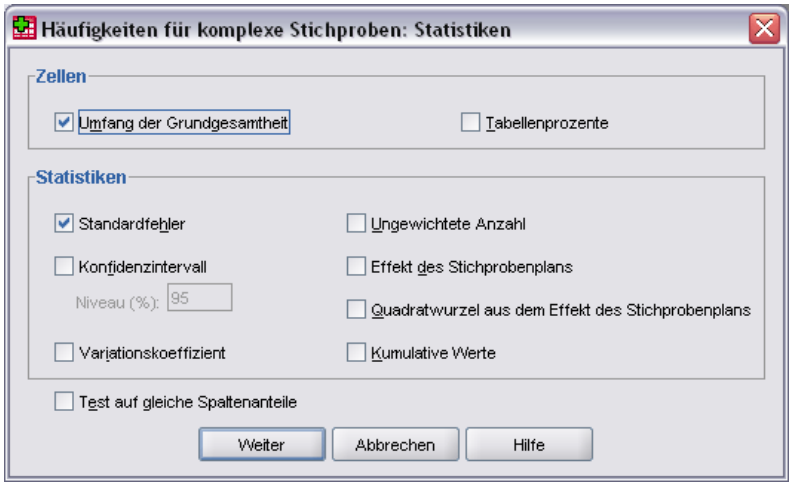


- Wählen Sie mindestens eine Häufigkeitsvariable aus.

Optional können Sie Variablen zur Definition von Teilgesamtheiten angeben. Die Statistiken werden für jede Teilgesamtheit getrennt berechnet.

Häufigkeiten für komplexe Stichproben: Statistiken

Abbildung 5-2
Dialogfeld "Häufigkeiten: Statistik"



Zellen. In dieser Gruppe können Sie Schätzungen für die Umfänge der Grundgesamtheiten für die Zellen und Tabellenprozentsätze abrufen.

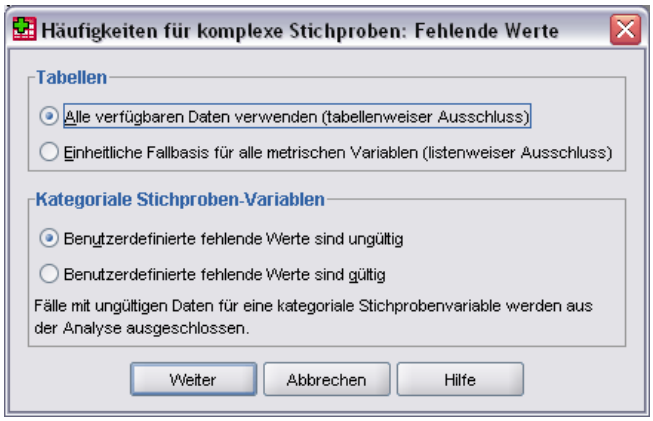
Statistiken. Mit dieser Gruppe können Statistiken zum Umfang der Grundgesamtheit oder zum Tabellenprozentsatz erstellt werden.

- **Standardfehler.** Der Standardfehler des Schätzers.
- **Konfidenzintervall.** Ein Konfidenzintervall für den Schätzer unter Verwendung der festgelegten Niveaus.
- **Variationskoeffizient.** Das Verhältnis des Standardfehlers des Schätzers zum Schätzer.
- **Ungewichtete Anzahl.** Die Anzahl der Einheiten, die zur Berechnung des Schätzers verwendet wurden.
- **Effekt des Stichprobenplans.** Das Verhältnis der Varianz des Schätzers zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Es handelt sich hierbei um ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.
- **Quadratwurzel aus dem Effekt des Stichprobenplans.** Dies ist ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.
- **Kumulative Werte.** Die kumulierte Schätzung über jeden einzelnen Wert der Variablen.

Test auf gleiche Spaltenanteile. Mit dieser Option werden Chi-Quadrat- und Likelihood-Quotienten-Tests für die Hypothese erstellt, dass die Kategorien einer Variable die gleichen Häufigkeiten aufweisen. Für jede Variable werden separate Tests durchgeführt.

Komplexe Stichproben: Fehlende Werte

Abbildung 5-3
Dialogfeld "Fehlende Werte"


Häufigkeiten für komplexe Stichproben: Fehlende Werte
Tabellen
Alle verfügbaren Daten verwenden (tabellenweiser Ausschluss)
Einheitliche Fallbasis für alle metrischen Variablen (listenweiser Ausschluss)
Kategoriale Stichproben-Variablen
Benutzerdefinierte fehlende Werte sind ungültig
Benutzerdefinierte fehlende Werte sind gültig
Fälle mit ungültigen Daten für eine kategoriale Stichprobenvariable werden aus der Analyse ausgeschlossen.
Weiter
Abbrechen
Hilfe

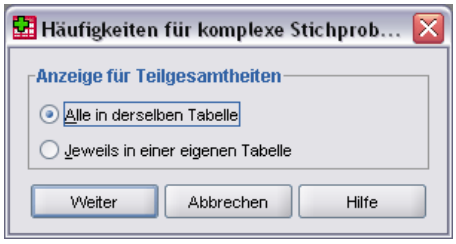
Tabellen. Diese Gruppe bestimmt, welche Fälle bei der Analyse verwendet werden.

- **Alle verfügbaren Daten verwenden.** Fehlende Werte werden jeweils für einzelne Tabellen bestimmt. Daher können die zur Berechnung von Statistiken verwendeten Fälle zwischen den einzelnen Häufigkeits- bzw. Kreuztabellen variieren.
- **Einheitliche Fallbasis verwenden.** Fehlende Werte werden über alle Variablen bestimmt. Daher sind die zur Berechnung der Statistiken verwendeten Fälle über die verschiedenen Tabellen konsistent.

Kategoriale Stichproben-Variablen. Diese Gruppe bestimmt, ob benutzerdefinierte fehlende Werte gültig oder ungültig sind.

Komplexe Stichproben: Optionen

Abbildung 5-4
Dialogfeld "Optionen"



Anzeige für Teilgesamtheiten. Sie können Teilgesamtheiten in derselben Tabelle oder in getrennten Tabellen anzeigen lassen.

Kapitel 6

Deskriptive Statistiken für komplexe Stichproben

Die Prozedur “Deskriptive Statistiken für komplexe Stichproben” zeigt univariate Auswertungsstatistiken für verschiedene Variablen an. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Beispiel. Mit der Prozedur “Deskriptive Statistiken für komplexe Stichproben” können Sie univariate deskriptive Statistiken für das Aktivitätsniveau von US-Bürgern erstellen, die auf den Ergebnissen der Umfrage National Health Interview Survey (NHIS) beruhen und einen geeigneten Analyseplan für diese öffentlich zugänglichen Daten beinhalten.

Statistiken. Mit diesem Verfahren erhalten Sie Mittelwerte und Summen sowie *T*-Tests, Standardfehler, Konfidenzintervalle, Variationskoeffizienten, die ungewichteten Anzahlen, den Umfang der Grundgesamtheiten, die Effekte des Stichprobenplans und die Quadratwurzeln aus den Effekten des Stichprobenplans für jede Schätzung.

Daten. Die Maße sollten metrische Variablen sein. Bei den Variablen für die Teilgesamtheiten kann es sich um String-Variablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Plan, deren Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld “Komplexe Stichproben: Plan” ausgewählt wurde.

Erstellen von deskriptiven Statistiken für komplexe Stichproben

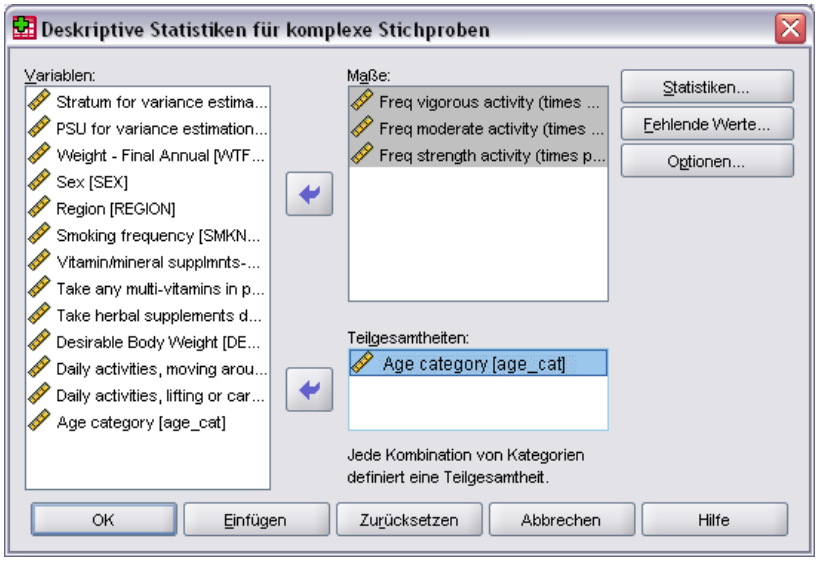
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Komplexe Stichproben
  Deskriptive Statistiken...

- Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.

- Klicken Sie auf Weiter.

Abbildung 6-1
Dialogfeld "Deskriptive Statistik"



▶ Wählen Sie mindestens eine metrische Variable aus.

Optional können Sie Variablen zur Definition von Teilgesamtheiten angeben. Die Statistiken werden für jede Teilgesamtheit getrennt berechnet.

Deskriptive Statistiken für komplexe Stichproben: Statistiken

Abbildung 6-2
Dialogfeld für deskriptive Statistiken

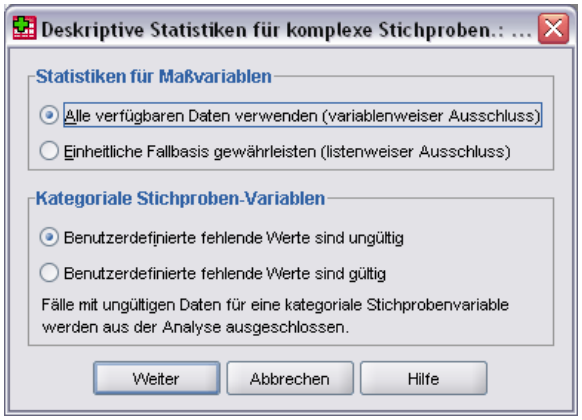
Auswertungen. Mit dieser Gruppe können Sie Schätzer der Mittelwerte und Summen der Maßvariablen abrufen. Außerdem können Sie *T*-Tests der Schätzer unter Verwendung eines angegebenen Werts abrufen.

Statistiken. Mit dieser Gruppe erhalten Sie Statistiken zu Mittelwert oder Summe.

- **Standardfehler.** Der Standardfehler des Schätzers.
- **Konfidenzintervall.** Ein Konfidenzintervall für den Schätzer unter Verwendung der festgelegten Niveaus.
- **Variationskoeffizient.** Das Verhältnis des Standardfehlers des Schätzers zum Schätzer.
- **Ungewichtete Anzahl.** Die Anzahl der Einheiten, die zur Berechnung des Schätzers verwendet wurden.
- **Umfang der Grundgesamtheit.** Die geschätzte Anzahl an Einheiten in der Grundgesamtheit.
- **Effekt des Stichprobenplans.** Das Verhältnis der Varianz des Schätzers zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Es handelt sich hierbei um ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.
- **Quadratwurzel aus dem Effekt des Stichprobenplans.** Dies ist ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.

Deskriptive Statistiken für komplexe Stichproben: Fehlende Werte

Abbildung 6-3
Dialogfeld für fehlende Werte bei deskriptiven Statistiken


Deskriptive Statistiken für komplexe Stichproben.: ...
Statistiken für Maßvariablen
Alle verfügbaren Daten verwenden (variablenweiser Ausschluss)
Einheitliche Fallbasis gewährleisten (listenweiser Ausschluss)
Kategoriale Stichproben-Variablen
Benutzerdefinierte fehlende Werte sind ungültig
Benutzerdefinierte fehlende Werte sind gültig
Fälle mit ungültigen Daten für eine kategoriale Stichprobenvariable werden aus der Analyse ausgeschlossen.
Weiter
Abbrechen
Hilfe

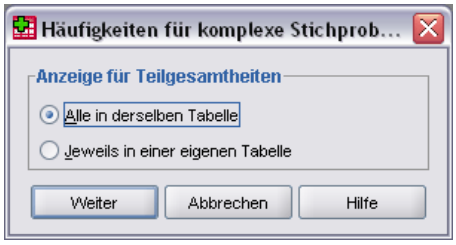
Statistiken für Maßvariablen. Diese Gruppe bestimmt, welche Fälle bei der Analyse verwendet werden.

- **Alle verfügbaren Daten verwenden.** Die fehlenden Werte werden variablenweise bestimmt, sodass die für die Berechnung der Statistiken verwendeten Werte bei den verschiedenen Maßvariablen voneinander abweichen können.
- **Einheitliche Fallbasis gewährleisten.** Die fehlenden Werte werden über alle Variablen bestimmt, sodass die für die Berechnung der Statistiken verwendeten Fälle konsistent sind.

Kategoriale Stichproben-Variablen. Diese Gruppe bestimmt, ob benutzerdefinierte fehlende Werte gültig oder ungültig sind.

Komplexe Stichproben: Optionen

Abbildung 6-4
Dialogfeld "Optionen"



Anzeige für Teilgesamtheiten. Sie können Teilgesamtheiten in derselben Tabelle oder in getrennten Tabellen anzeigen lassen.

Kapitel
7

Kreuztabellen für komplexe Stichproben

Mit der Prozedur “Kreuztabellen für komplexe Stichproben” werden Kreuztabellen für Paare von ausgewählten Variablen erstellt und bivariate Statistiken angezeigt. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Beispiel. Mit der Prozedur “Kreuztabellen für komplexe Stichproben” können Sie Kreuzklassifikationsstatistiken für die Häufigkeit des Rauchens im Verhältnis zur Einnahme von Vitaminpräparaten bei US-Bürgern erstellen, die auf den Ergebnissen der Umfrage National Health Interview Survey (NHIS) beruhen und einen geeigneten Analyseplan für diese öffentlich zugänglichen Daten beinhalten.

Statistiken. Mit diesem Verfahren erhalten Sie Schätzungen für die Umfänge der Grundgesamtheiten der Zellen sowie Prozentsätze für Zeilen, Spalten und die Tabelle, außerdem Standardfehler, Konfidenzintervalle, Variationskoeffizienten, erwartete Werte, Effekte des Stichprobenplans, Quadratwurzeln aus den Effekten des Stichprobenplans, Residuen, korrigierte Residuen sowie die ungewichtete Anzahl für jede Schätzung. Das Quotenverhältnis, das relative Risiko und die Risiko-Differenz werden für 2x2-Tabellen berechnet. Des Weiteren werden die Pearson-Statistik und die Statistik für den Likelihood-Quotienten für den Test auf Unabhängigkeit der Zeilen- und Spaltenvariablen berechnet.

Daten. Die Zeilen- und Spaltenvariablen sollten kategorial sein. Bei den Variablen für die Teilgesamtheiten kann es sich um String-Variablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Plan, deren Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld “Komplexe Stichproben: Plan” ausgewählt wurde.

Erstellen von Kreuztabellen für komplexe Stichproben

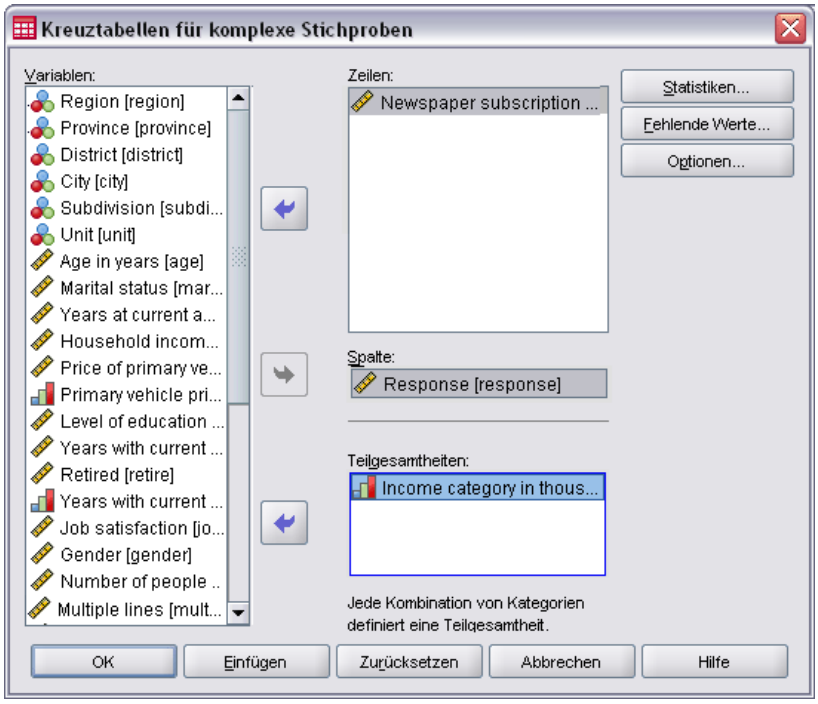
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Komplexe Stichproben
  Kreuztabellen...

- Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.

- Klicken Sie auf Weiter.

Abbildung 7-1
Dialogfeld "Kreuztabellen"



▶ Wählen Sie mindestens eine Zeilen- und eine Spaltenvariable aus.

Optional können Sie Variablen zur Definition von Teilgesamtheiten angeben. Die Statistiken werden für jede Teilgesamtheit getrennt berechnet.

Kreuztabellen für komplexe Stichproben - Statistik

Abbildung 7-2
Dialogfeld "Kreuztabellen: Statistik"

Zellen. In dieser Gruppe können Sie Schätzungen für den Umfang der Grundgesamtheit für die Zellen sowie Prozentsätze für Zeilen, Spalten und die Tabelle abrufen.

Statistiken. In dieser Gruppe finden Sie Statistiken zu den Umfängen der Grundgesamtheiten und zu den Prozentsätzen für Zeilen, Spalten und die Tabelle.

- **Standardfehler.** Der Standardfehler des Schätzers.
- **Konfidenzintervall.** Ein Konfidenzintervall für den Schätzer unter Verwendung der festgelegten Niveaus.
- **Variationskoeffizient.** Das Verhältnis des Standardfehlers des Schätzers zum Schätzer.
- **Erwartete Werte.** Der erwartete Wert des Schätzers, unter der Hypothese der Unabhängigkeit der Zeilen- und Spaltenvariable.
- **Ungewichtete Anzahl.** Die Anzahl der Einheiten, die zur Berechnung des Schätzers verwendet wurden.
- **Effekt des Stichprobenplans.** Das Verhältnis der Varianz des Schätzers zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Es handelt sich hierbei um ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.
- **Quadratwurzel aus dem Effekt des Stichprobenplans.** Dies ist ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.

- **Residuen.** Der erwartete Wert ist die Anzahl von Fällen, die in einer Zelle erwartet würden, wenn kein Zusammenhang zwischen den beiden Variablen bestünde. Ein positives Residuum zeigt an, dass in der Zelle mehr Fälle vorliegen, als dies der Fall wäre, wenn die Zeilen- und Spaltenvariable unabhängig wären.
- **Korrigierte Residuen.** Der Quotient aus dem Residuum einer Zelle (beobachteter Wert minus erwarteter Wert) und dessen geschätztem Standardfehler. Das resultierende standardisierte Residuum wird in Einheiten der Standardabweichung über oder unter dem Mittelwert angegeben.

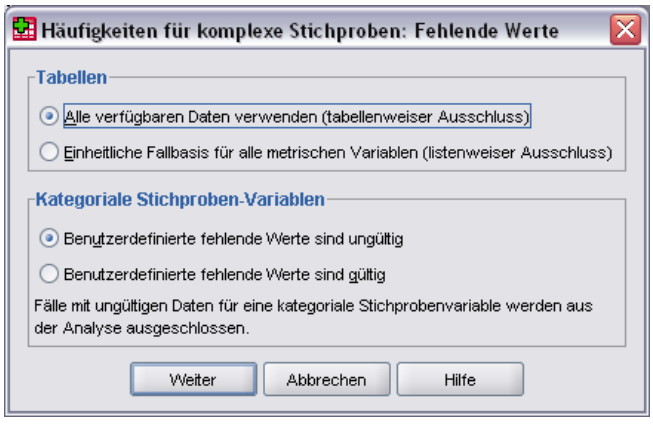
Auswertungen für 2x2-Tabellen. In dieser Gruppe finden Sie Statistiken für Tabellen, in denen die Zeilen- und die Spaltenvariable jeweils zwei Kategorien aufweisen. Beide messen die Stärke des Zusammenhangs zwischen dem Vorhandensein eines Faktors und dem Auftreten eines Ereignisses.

- **Quotenverhältnis.** Das Quotenverhältnis kann als Schätzer des relativen Risikos verwendet werden, wenn der Faktor selten auftritt.
- **Relatives Risiko.** Das Verhältnis zwischen dem Risiko eines Ereignisses bei Vorliegen des Faktors zum Risiko des Ereignisses bei Fehlen des Faktors.
- **Risiko-Differenz.** Die Differenz zwischen dem Risiko eines Ereignisses bei Vorliegen des Faktors zum Risiko des Ereignisses bei Fehlen des Faktors.

Test auf Unhabhängigkeit von Zeilen und Spalten. Mit dieser Option werden Chi-Quadrat- und Likelihood-Quotienten-Tests für die Hypothese erstellt, dass eine Zeilen- und eine Spaltenvariable unabhängig sind. Für jedes Variablenpaar werden separate Tests durchgeführt.

Komplexe Stichproben: Fehlende Werte

Abbildung 7-3
Dialogfeld "Fehlende Werte"


Häufigkeiten für komplexe Stichproben: Fehlende Werte
Tabellen
Alle verfügbaren Daten verwenden (tabellenweiser Ausschluss)
Einheitliche Fallbasis für alle metrischen Variablen (listenweiser Ausschluss)
Kategoriale Stichproben-Variablen
Benutzerdefinierte fehlende Werte sind ungültig
Benutzerdefinierte fehlende Werte sind gültig
Fälle mit ungültigen Daten für eine kategoriale Stichprobenvariable werden aus der Analyse ausgeschlossen.
Weiter
Abbrechen
Hilfe

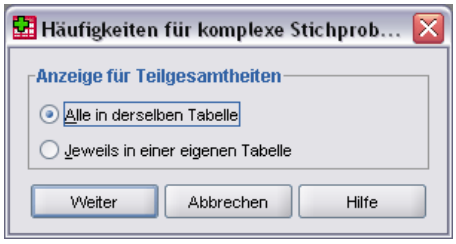
Tabellen. Diese Gruppe bestimmt, welche Fälle bei der Analyse verwendet werden.

- **Alle verfügbaren Daten verwenden.** Fehlende Werte werden jeweils für einzelne Tabellen bestimmt. Daher können die zur Berechnung von Statistiken verwendeten Fälle zwischen den einzelnen Häufigkeits- bzw. Kreuztabellen variieren.
- **Einheitliche Fallbasis verwenden.** Fehlende Werte werden über alle Variablen bestimmt. Daher sind die zur Berechnung der Statistiken verwendeten Fälle über die verschiedenen Tabellen konsistent.

Kategoriale Stichproben-Variablen. Diese Gruppe bestimmt, ob benutzerdefinierte fehlende Werte gültig oder ungültig sind.

Komplexe Stichproben: Optionen

Abbildung 7-4
Dialogfeld "Optionen"



Anzeige für Teilgesamtheiten. Sie können Teilgesamtheiten in derselben Tabelle oder in getrennten Tabellen anzeigen lassen.

Kapitel 8

Verhältnisse für komplexe Stichproben

Die Prozedur "Verhältnisse für komplexe Stichproben" zeigt univariate Auswertungsstatistiken für Verhältnisse von Variablen an. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Beispiel. Mit der Prozedur "Verhältnisse für komplexe Stichproben" können Sie deskriptive Statistiken für das Verhältnis des aktuellen Eigenschaftswerts zum letzten bewerteten Wert erstellen. Diese Statistiken beruhen auf den Ergebnissen einer Studie, die anhand eines komplexen Plans und mit einem geeigneten Analyseplan für die Daten in einem US-Bundesstaat durchgeführt wurde.

Statistiken. Mit diesem Verfahren erhalten Sie Verhältnisschätzer, *T*-Tests, Standardfehler, Konfidenzintervalle, Variationskoeffizienten, die ungewichteten Anzahlen, den Umfang der Grundgesamtheiten, die Effekte des Stichprobenplans und die Quadratwurzeln aus den Effekten des Stichprobenplans.

Daten. Zähler und Nenner sollten metrische Variablen mit positivem Wert sein. Bei den Variablen für die Teilgesamtheiten kann es sich um String-Variablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Plan, deren Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld "Komplexe Stichproben: Plan" ausgewählt wurde.

Ermitteln von Verhältnissen für komplexe Stichproben

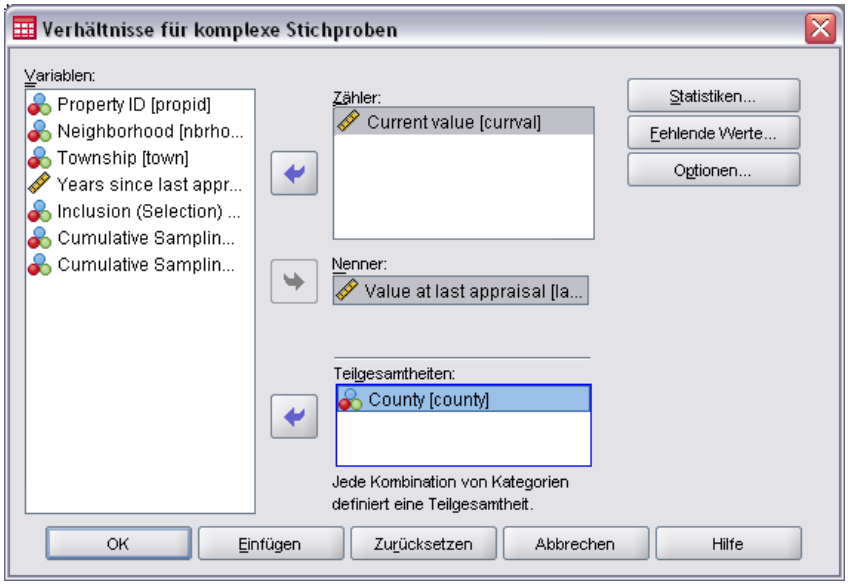
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Komplexe Stichproben
  Verhältnisse...

- Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.

- Klicken Sie auf Weiter.

Abbildung 8-1
Dialogfeld "Verhältnisse"

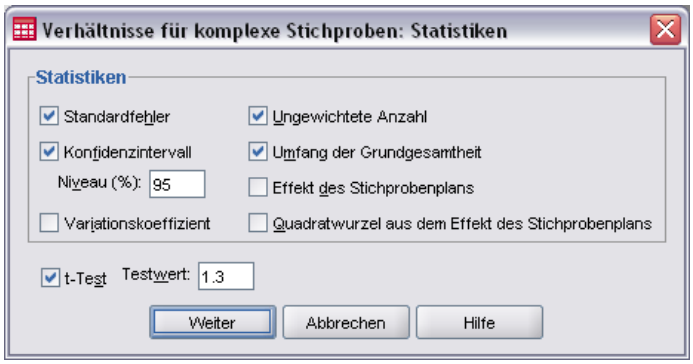


▶ Wählen Sie mindestens eine Zähler- und eine Nennervariable aus.

Optional können Sie Variablen zur Definition der Untergruppen angeben, für die Statistiken erstellt werden.

Verhältnisse für komplexe Stichproben: Statistiken

Abbildung 8-2
Dialogfeld "Verhältnisse: Statistiken"



Statistiken. Mit dieser Gruppe erhalten Sie Statistiken zum Verhältnisschätzer.

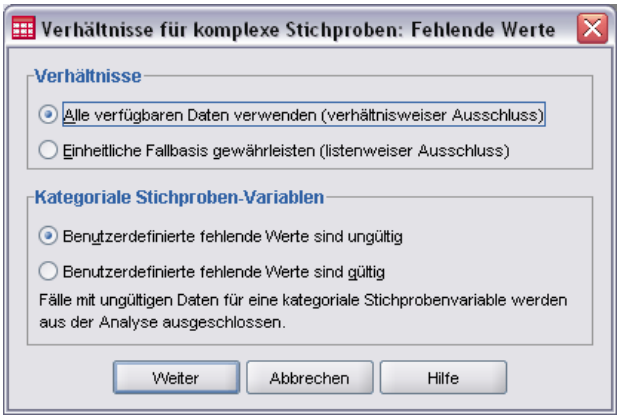
- **Standardfehler.** Der Standardfehler des Schätzers.
- **Konfidenzintervall.** Ein Konfidenzintervall für den Schätzer unter Verwendung der festgelegten Niveaus.
- **Variationskoeffizient.** Das Verhältnis des Standardfehlers des Schätzers zum Schätzer.
- **Ungewichtete Anzahl.** Die Anzahl der Einheiten, die zur Berechnung des Schätzers verwendet wurden.

- **Umfang der Grundgesamtheit.** Die geschätzte Anzahl an Einheiten in der Grundgesamtheit.
- **Effekt des Stichprobenplans.** Das Verhältnis der Varianz des Schätzers zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Es handelt sich hierbei um ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.
- **Quadratwurzel aus dem Effekt des Stichprobenplans.** Dies ist ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.

T-Test. Sie können *T*-Tests der Schätzer unter Verwendung eines angegebenen Werts abrufen.

Verhältnisse für komplexe Stichproben: Fehlende Werte

Abbildung 8-3
Dialogfeld "Verhältnisse: Fehlende Werte"


Verhältnisse für komplexe Stichproben: Fehlende Werte
Verhältnisse
Alle verfügbaren Daten verwenden (verhältnisweiser Ausschluss)
Einheitliche Fallbasis gewährleisten (listenweiser Ausschluss)
Kategoriale Stichproben-Variablen
Benutzerdefinierte fehlende Werte sind ungültig
Benutzerdefinierte fehlende Werte sind gültig
Fälle mit ungültigen Daten für eine kategoriale Stichprobenvariable werden aus der Analyse ausgeschlossen.
Weiter
Abbrechen
Hilfe

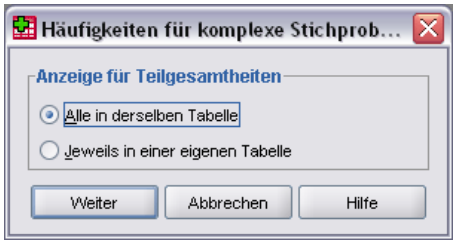
Verhältnisse. Diese Gruppe bestimmt, welche Fälle bei der Analyse verwendet werden.

- **Alle verfügbaren Daten verwenden.** Fehlende Werte werden jeweils für einzelne Verhältnisse bestimmt. Daher können die zur Berechnung von Statistiken verwendeten Fälle zwischen den einzelnen Zähler/Nenner-Paaren variieren.
- **Einheitliche Fallbasis gewährleisten.** Fehlende Werte werden über alle Variablen bestimmt. Daher sind die zur Berechnung der Statistiken verwendeten Fälle konsistent.

Kategoriale Stichproben-Variablen. Diese Gruppe bestimmt, ob benutzerdefinierte fehlende Werte gültig oder ungültig sind.

Komplexe Stichproben: Optionen

Abbildung 8-4
Dialogfeld "Optionen"



Anzeige für Teilgesamtheiten. Sie können Teilgesamtheiten in derselben Tabelle oder in getrennten Tabellen anzeigen lassen.

Kapitel
9

Allgemeines lineares Modell für komplexe Stichproben

Die Prozedur “Allgemeines lineares Modell für komplexe Stichproben” besteht aus einer linearen Regressionsanalyse sowie aus einer Analyse der Varianz und Kovarianz für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Optional können Sie auch Analysen für eine Teilgesamtheit vornehmen.

Beispiel. Eine Lebensmittelkette hat eine Kundenumfrage über die Kaufgewohnheiten durchgeführt, die nach einem komplexen Plan ausgeführt wurde. Auf der Grundlage der Umfrageergebnisse und der Zahlen über die Ausgaben der einzelnen Kunden im vergangenen Monat möchte das Unternehmen ermitteln, ob die Einkaufshäufigkeit in einem Zusammenhang mit den monatlichen Ausgaben steht, und zwar getrennt nach Geschlecht. Bei dieser Untersuchung soll der Stichprobenplan berücksichtigt werden.

Statistiken. Mit dieser Prozedur erhalten Sie Schätzungen, Standardfehler, Konfidenzintervalle, *t*-Tests, Effekte des Stichprobenplans und Quadratwurzeln aus den Effekten des Stichprobenplans, außerdem die Korrelationen und Kovarianzen bei den Parameterschätzern. Auch Maße für die Anpassungsgüte des Modells und deskriptive Statistken für die abhängigen und unabhängigen Variablen stehen zur Verfügung. Und nicht zuletzt können Sie geschätzte Randmittel für die Modellfaktorebenen und die Wechselwirkungen zwischen den Faktoren anfordern.

Daten. Die abhängige Variable ist quantitativ. Faktoren sind kategorial. Kovariaten sind quantitative Variablen, die mit der abhängigen Variablen in Beziehung stehen. Bei den Variablen für die Teilgesamtheiten kann es sich um String-Variablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Plan, deren Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld “Komplexe Stichproben: Plan” ausgewählt wurde.

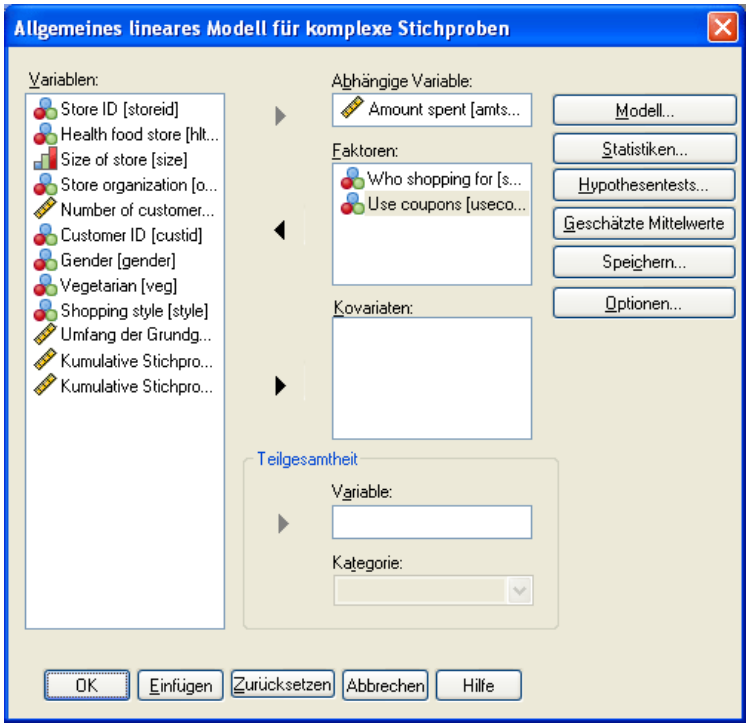
Erzeugen eines allgemeinen linearen Modells für komplexe Stichproben

Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Komplexe Stichproben
    Allgemeines lineares Modell...

- Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.
- Klicken Sie auf Weiter.

Abbildung 9-1
Dialogfeld "Allgemeines lineares Modell"

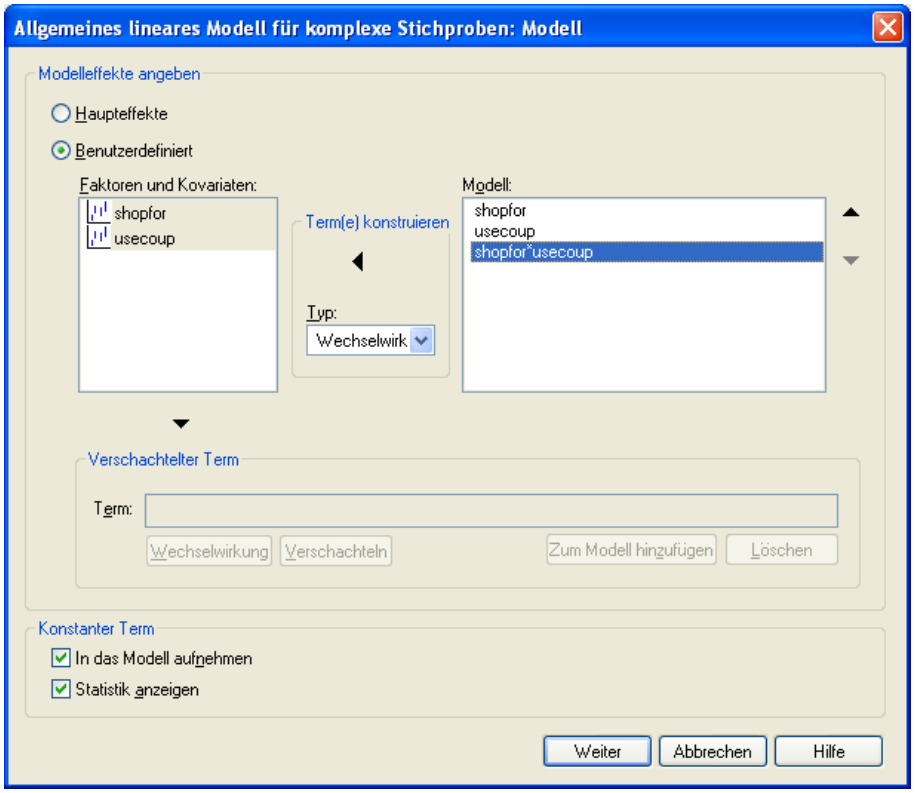


- Wählen Sie eine abhängige Variable aus.

Die folgenden Optionen sind verfügbar:

- Wählen Sie die entsprechenden Variablen für Faktoren und Kovariaten gemäß den vorliegenden Daten aus.
- Legen Sie eine Variable fest, um so eine Teilgesamtheit zu definieren. Die Analyse wird ausschließlich für die ausgewählte Kategorie der Teilgesamtheit-Variable vorgenommen.

Abbildung 9-2
Dialogfeld "Modell"



Modell-Effekte angeben. Standardmäßig erstellt das Verfahren ein Modell mit Haupteffekten unter Verwendung der im Hauptdialogfeld angegebenen Faktoren und Kovariaten. Alternativ können Sie ein benutzerdefiniertes Modell erstellen, das Wechselwirkungseffekte und verschachtelte Terme enthält.

Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung für alle ausgewählten Variablen erzeugt.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-fach. Hiermit werden alle möglichen 2-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-fach. Hiermit werden alle möglichen 3-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-fach. Hiermit werden alle möglichen 4-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-fach. Hiermit werden alle möglichen 5-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Verschachtelte Terme

In dieser Prozedur können Sie verschachtelte Terme für ein Modell konstruieren. Verschachtelte Terme sind nützlich, um den Effekt von Faktoren oder Kovariaten zu analysieren, deren Werte nicht mit den Stufen eines anderen Faktors interagieren. Eine Lebensmittelkette kann beispielsweise das Kaufverhalten ihrer Kunden in mehreren Filialen untersuchen. Da jeder Kunde nur eine dieser Filialen besucht, kann der Effekt *Kunde* als **verschachtelt innerhalb** des Effekts *Filiale* beschrieben werden.

Darüber hinaus können Sie Wechselwirkungseffekte, wie polynomiale Terme mit derselben Kovariaten, einschließen oder dem verschachtelten Term mehrere Verschachtelungsebenen hinzufügen.

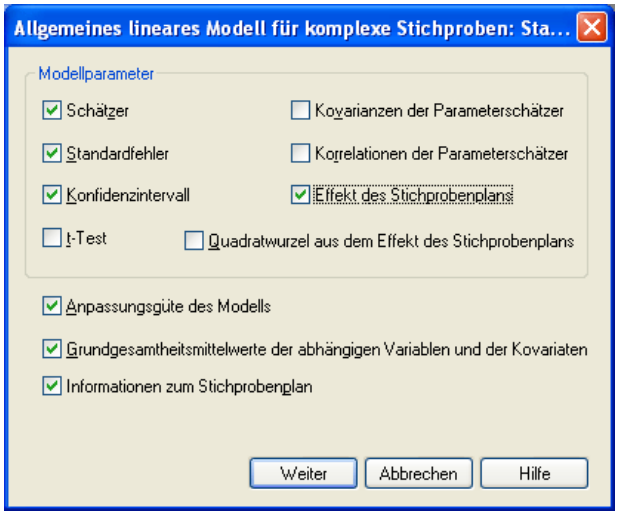
Einschränkungen. Für verschachtelte Terme gelten die folgenden Einschränkungen:

- Alle Faktoren innerhalb einer Wechselwirkung müssen eindeutig sein. Dementsprechend ist die Angabe von *A*A* unzulässig, wenn *A* ein Faktor ist.
- Alle Faktoren innerhalb eines verschachtelten Effekts müssen eindeutig sein. Dementsprechend ist die Angabe von *A(A)* unzulässig, wenn *A* ein Faktor ist.
- Effekte dürfen nicht in einer Kovariaten verschachtelt werden. Dementsprechend ist die Angabe von *A(X)* unzulässig, wenn *A* ein Faktor und *X* eine Kovariate ist.

Konstanter Term. Der konstante Term wird gewöhnlich in das Modell aufgenommen. Wenn anzunehmen ist, dass die Daten durch den Koordinatenursprung verlaufen, können Sie den konstanten Term ausschließen. Selbst wenn Sie den konstanten Term in das Modell aufnehmen, können Sie festlegen, dass die darauf bezogenen Statistiken unterdrückt werden sollen.

Allgemeines lineares Modell für komplexe Stichproben: Statistik

Abbildung 9-3
Dialogfeld "Allgemeines lineares Modell: Statistik"



Modellparameter. In dieser Gruppe steuern Sie die Anzeige der Statistiken für die Modellparameter.

- **Schätzer.** Zeigt eine Schätzung der Koeffizienten.

- **Standardfehler.** Zeigt den Standardfehler für die einzelnen Koeffizientenschätzer.
- **Konfidenzintervall.** Zeigt ein Konfidenzintervall für die einzelnen Koeffizientenschätzer. Das Konfidenzniveau für das Intervall wird im Dialogfeld "Optionen" festgelegt.
- **T-Test.** Zeigt je einen *t*-Test für die einzelnen Koeffizientenschätzer. Die Nullhypothese der Tests liegt vor, wenn der Koeffizient den Wert 0 aufweist.
- **Kovarianzen der Parameterschätzer.** Zeigt eine Schätzung der Kovarianzmatrix für die Modellkoeffizienten.
- **Korrelationen der Parameterschätzer.** Zeigt eine Schätzung der Korrelationsmatrix für die Modellkoeffizienten.
- **Effekt des Stichprobenplans.** Das Verhältnis der Varianz des Schätzers zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Es handelt sich hierbei um ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.
- **Quadratwurzel aus dem Effekt des Stichprobenplans.** Dies ist ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.

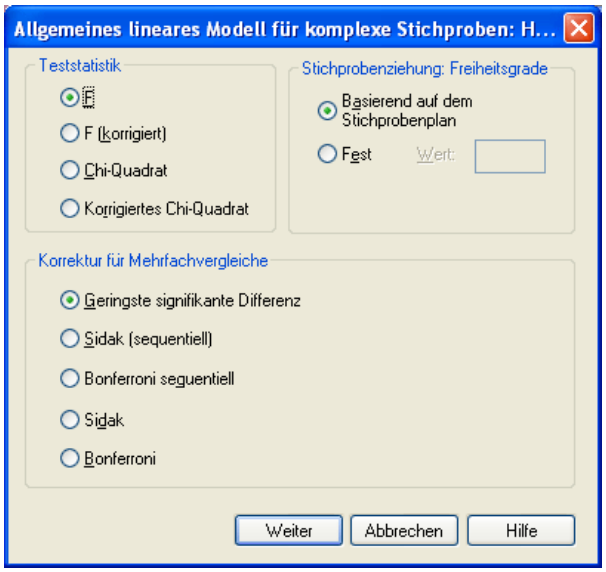
Anpassungsgüte des Modells. Zeigt die Statistik für R^2-Fehler und mittlere quadratische Fehler.

Mittelwerte der Grundgesamtheit für die abhängige Variable und die Kovariaten. Zeigt eine Zusammenfassung über die abhängige Variable, die Kovariaten und die Faktoren.

Informationen zum Stichprobenplan. Zeigt eine Zusammenfassung über die Stichprobe (mit ungewichteter Anzahl und Umfang der Grundgesamtheit).

Hypothesentests für komplexe Stichproben

Abbildung 9-4
Dialogfeld "Hypothesentests"



Teststatistik. In dieser Gruppe können Sie den Typ der Statistik zum Testen der Hypothesen festlegen. Die folgenden Optionen stehen zur Auswahl: *F*, *F* (korrigiert), "Chi-Quadrat" und "Korrigiertes Chi-Quadrat".

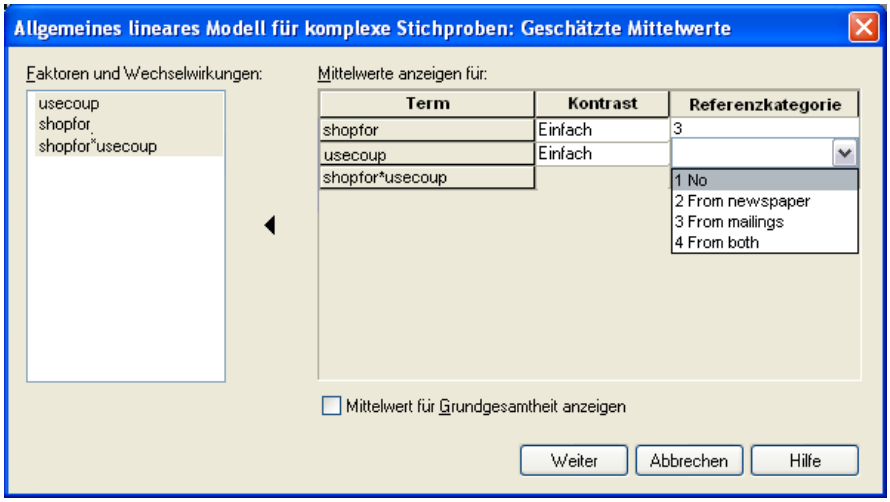
Stichprobenziehung: Freiheitsgrade. In dieser Gruppe steuern Sie die Freiheitsgrade im Stichprobenplan, mit denen die *p*-Werte für alle Teststatistiken berechnet werden. Dient der Stichprobenplan als Grundlage, ist dieser Wert die Differenz zwischen der Anzahl der primären Stichprobeneinheiten und der Anzahl der Schichten in der ersten Stufe der Stichproben. Alternativ können Sie benutzerdefinierte Freiheitsgrade festlegen; geben Sie hierzu eine positive Ganzzahl ein.

Korrektur für Mehrfachvergleiche. Bei der Durchführung von Hypothesentests mit mehreren Kontrasten kann das Gesamtsignifikanzniveau mithilfe der Signifikanzniveaus der eingeschlossenen Kontraste angepasst werden. In dieser Gruppe können Sie die Anpassungs-/Korrekturmethode auswählen.

- **Geringste signifikante Differenz.** Diese Methode steuert nicht die Gesamtwahrscheinlichkeit, dass Hypothesen abgelehnt werden, bei denen einige lineare Kontraste von den Werten einer Nullhypothese abweichen.
- **Sidak (sequenziell).** Hierbei handelt es sich um ein sequenzielles schrittweises Sidak-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen anbelangt, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.
- **Bonferroni sequenziell.** Hierbei handelt es sich um ein sequenzielles schrittweises Bonferroni-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen anbelangt, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.
- **Sidak.** Dieses Verfahren liefert engere Grenzen als der Bonferroni-Ansatz.
- **Bonferroni.** Dieses Verfahren passt das empirische Signifikanzniveau der Tatsache an, dass mehrere Kontraste getestet werden.

Allgemeines lineares Modell für komplexe Stichproben: Geschätzte Mittelwerte

Abbildung 9-5
Dialogfeld "Allgemeines lineares Modell: Geschätzte Mittelwerte."



Im Dialogfeld "Geschätzte Mittelwerte" werden die vom Modell geschätzten Randmittel für die Ebenen der Faktoren und die Wechselwirkungen zwischen Faktoren aufgeführt, die im untergeordneten Dialogfeld "Modell" angegeben wurden. Des Weiteren können Sie den Mittelwert für die gesamte Grundgesamtheit anzeigen lassen.

Term. Geschätzte Mittel werden für die ausgewählten Faktoren und Wechselwirkungen zwischen Faktoren berechnet.

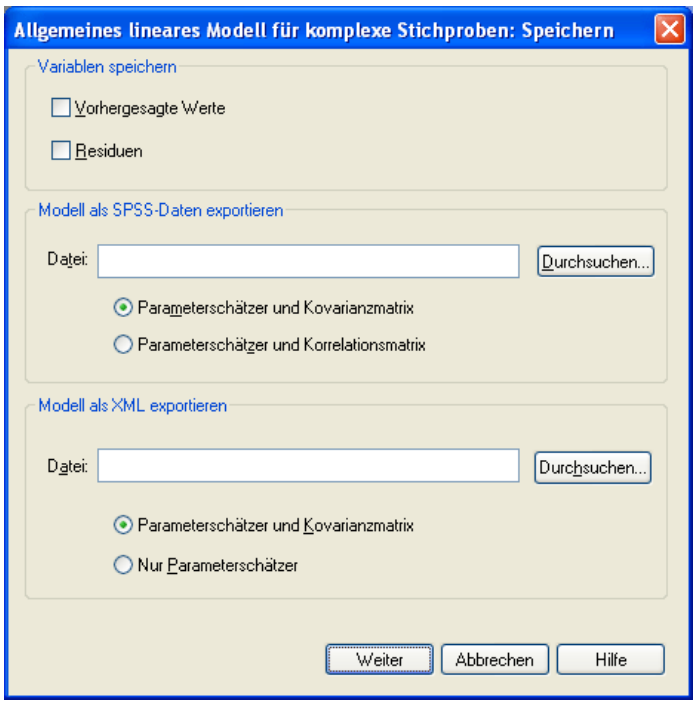
Kontrast. Der Kontrast bestimmt, wie die Hypothesentests zum Vergleich der geschätzten Mittel eingerichtet werden.

- **Einfach.** Vergleicht den Mittelwert jeder Stufe mit dem Mittelwert einer vorgegebenen Stufe. Diese Art von Kontrast ist nützlich, wenn es eine Kontrollgruppe gibt.
- **Abweichung.** Vergleicht den Mittelwert jeder Stufe (mit Ausnahme einer Referenzkategorie) mit dem Mittelwert aller Stufen (Gesamtmittel). Die Stufen des Faktors können in beliebiger Reihenfolge stehen.
- **Differenz.** Vergleicht den Mittelwert jeder Stufe (mit Ausnahme der ersten) mit dem Mittelwert der vorangehenden Stufen. Diese Kontraste werden auch als umgekehrte Helmert-Kontraste bezeichnet.
- **Helmert.** Vergleicht den Mittelwert jeder Stufe des Faktors (mit Ausnahme der letzten) mit dem Mittelwert der folgenden Stufen.
- **Wiederholt.** Vergleicht den Mittelwert jeder Stufe (außer der letzten) mit dem Mittelwert der folgenden Stufe.
- **Polynomial.** Vergleicht den linearen, quadratischen, kubischen Effekt usw. Der erste Freiheitsgrad enthält den linearen Effekt über alle Kategorien; der zweite Freiheitsgrad den quadratischen Effekt usw. Diese Kontraste werden häufig verwendet, um polynomiale Trends zu schätzen.

Referenzkategorie. Für einfache Kontraste und Abweichungskontraste wird eine Referenzkategorie oder eine Faktorenebene benötigt, mit der die anderen Elemente verglichen werden können.

Allgemeines lineares Modell für komplexe Stichproben: Speichern

Abbildung 9-6
Dialogfeld "Allgemeines lineares Modell: Speichern"



Variablen speichern. In dieser Gruppe speichern Sie die vom Modell vorhergesagten Werte und Residuen als neue Variablen in der Arbeitsdatei.

Modell als SPSS-Daten exportieren. Schreibt ein SPSS-Daten-Set, das die Parameter-Korrelations- oder -Kovarianzmatrix enthält (mit Parameterschätzern, Standardfehlern, Signifikanzwerten und Freiheitsgraden). Die Reihenfolge der Variablen in der Matrixdatei lautet wie folgt:

- **rowtype_.** Nimmt folgende Werte (und Wertelabel) an: COV (Kovarianzen), CORR (Korrelationen), EST (Parameterschätzer), SE (Standardfehler), SIG (Signifikanzniveaus) und DF (Freiheitsgrade des Stichprobenplans). Es gibt einen separaten Fall mit dem Zeilentyp COV (bzw. CORR) für jeden Modellparameter sowie einen separaten Fall für jeden der anderen Zeilentypen.
- **varname_.** Nimmt für die Zeilentypen COV bzw. CORR die Werte P1, P2, ... an, was einer geordneten Liste aller geschätzten Modellparameter entspricht, mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Für alle anderen Zeilentypen sind die Zellen leer.
- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Die Werte hängen jeweils vom Zeilentyp ab. Bei redundanten Parametern sind

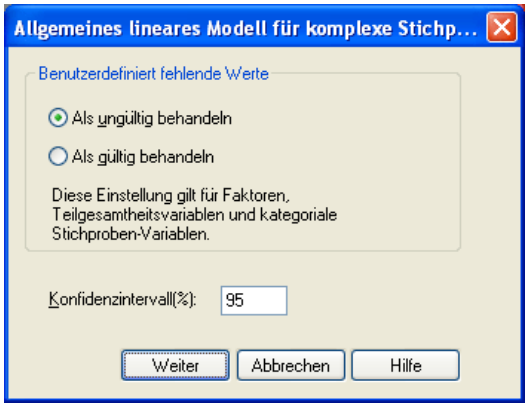
alle Kovarianzen auf 0 gesetzt, die Korrelationen sind auf den systemdefiniert fehlenden Wert gesetzt, alle Parameterschätzer sind auf 0 gesetzt und alle Standardfehler, Signifikanzniveaus und die Freiheitsgrade der Residuen sind auf den systemdefiniert fehlenden Wert gesetzt.

Anmerkung: Diese Datei ist nicht unmittelbar für weitere Analysen in anderen Prozeduren verwendbar, bei denen eine Matrixdatei eingelesen wird, es sei denn, diese Prozeduren akzeptieren alle hier exportierten Zeilentypen.

Modell als XML exportieren. Speichert die Parameterschätzungen und ggf. die Parameter-Kovarianzmatrix (falls ausgewählt) im XML-Format (PMML). SmartScore und SPSS Server (gesondertes Produkt) können anhand dieser Modelldatei die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.

Allgemeines lineares Modell für komplexe Stichproben: Optionen

Abbildung 9-7
Dialogfeld "Allgemeines lineares Modell: Optionen"



Benutzerdefinierte fehlende Werte. Alle Stichproben-Variablen sowie die abhängige Variable und ggf. alle Kovariaten müssen gültige Daten enthalten. Fälle, bei denen ungültige Daten für diese Variablen vorliegen, werden aus der Analyse gelöscht. Mit diesen Steuerungen legen Sie fest, ob benutzerdefiniert fehlende Werte bei den Schicht-, Cluster-, Teilgesamtheits- und Faktorvariablen als gültige Werte behandelt werden sollen.

Konfidenzintervall. Dies ist die Konfidenzintervall-Ebene für Koeffizientenschätzungen und geschätzte Randmittel. Geben Sie einen Wert größer oder gleich 50 und kleiner als 100 ein.

Zusätzliche Funktionen beim Befehl CSGLM

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Mit dem Unterbefehl `CUSTOM` können Sie benutzerdefinierte Tests auf Effekte im Vergleich zu linearen Kombinationen von Effekten oder einem Wert vornehmen.
- Mit dem Unterbefehl `EMMEANS` können Sie bei der Berechnung der geschätzten Randmittel einen anderen Wert für die Kovariaten festlegen als den Mittelwert.
- Mit dem Unterbefehl `EMMEANS` können Sie bei polynomialen Kontrasten eine Metrik angeben.

- Mit dem Unterbefehl `CRITERIA` können Sie einen Toleranzwert für die Prüfung auf Singularität festlegen.
- Mit dem Unterbefehl `SAVE` können Sie benutzerdefinierte Namen für gespeicherte Variablen angeben.
- Mit dem Unterbefehl `PRINT` können Sie eine Tabelle mit allgemeinen schätzbaren Funktionen anlegen.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
10

Logistische Regression für komplexe Stichproben

Die Prozedur "Logistische Regression für komplexe Stichproben" besteht aus einer logistischen Regressionsanalyse einer binären oder multinomialen abhängigen Variable für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Optional können Sie auch Analysen für eine Teilgesamtheit vornehmen.

Beispiel. Ein Kreditsachbearbeiter verfügt über eine Reihe von Datensätzen zu Kunden, die ein Darlehen in verschiedenen Zweigstellen erhalten haben; diese Datensätze wurden nach einem komplexen Plan zusammengestellt. Bei der Einbeziehung des Stichprobenplans interessiert sich der Sachbearbeiter für die Wahrscheinlichkeit, mit der ein Kunde mit dem Darlehen in Verzug geraten könnte, und zwar im Zusammenhang mit dem Alter, der beruflichen Entwicklung und der Darlehenshöhe.

Statistiken. Mit dieser Prozedur erhalten Sie Schätzungen, potenzierte Schätzungen, Standardfehler, Konfidenzintervalle, *t*-Tests, Effekte des Stichprobenplans und Quadratwurzeln aus den Effekten des Stichprobenplans, außerdem die Korrelationen und Kovarianzen bei den Parameterschätzern. Auch Pseudo-R^2-Statistiken, Klassifizierungstabellen und deskriptive Statistiken für die abhängigen und unabhängigen Variablen stehen zur Verfügung.

Daten. Die abhängige Variable ist kategorial. Faktoren sind kategorial. Kovariaten sind quantitative Variablen, die mit der abhängigen Variablen in Beziehung stehen. Bei den Variablen für die Teilgesamtheiten kann es sich um String-Variablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Plan, deren Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld "Komplexe Stichproben: Plan" ausgewählt wurde.

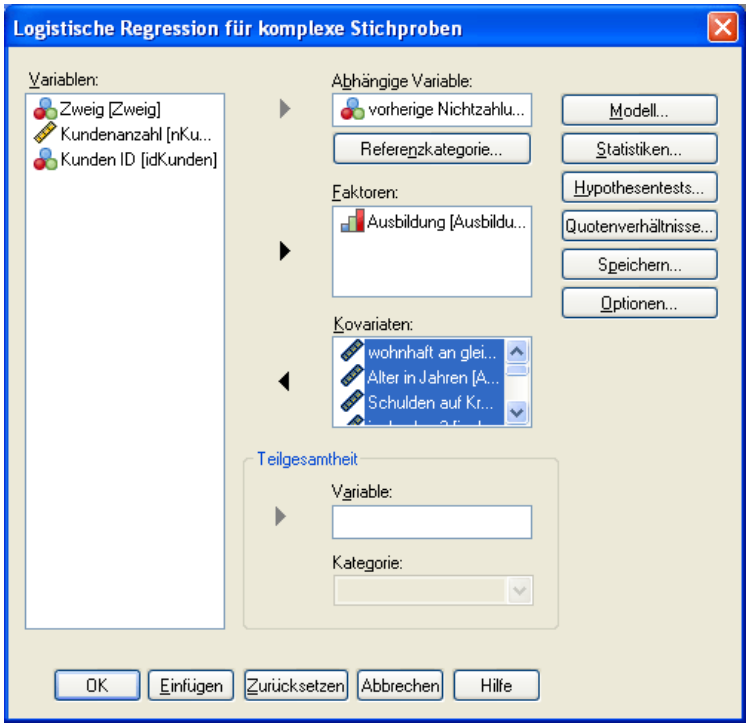
Ermitteln der logistischen Regression für komplexe Stichproben

Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Komplexe Stichproben
    Logistische Regression...

- ▶ Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.
- ▶ Klicken Sie auf Weiter.

Abbildung 10-1
Dialogfeld "Logistische Regression"



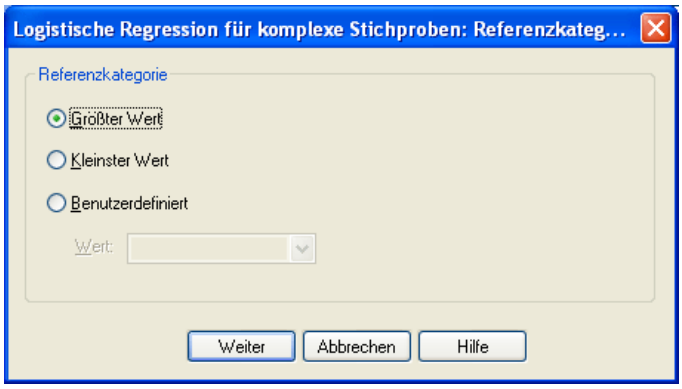
▶ Wählen Sie eine abhängige Variable aus.

Die folgenden Optionen sind verfügbar:

- Wählen Sie die entsprechenden Variablen für Faktoren und Kovariaten gemäß den vorliegenden Daten aus.
- Legen Sie eine Variable fest, um so eine Teilgesamtheit zu definieren. Die Analyse wird ausschließlich für die ausgewählte Kategorie der Teilgesamtheit-Variable vorgenommen.

Logistische Regression für komplexe Stichproben: Referenzkategorie

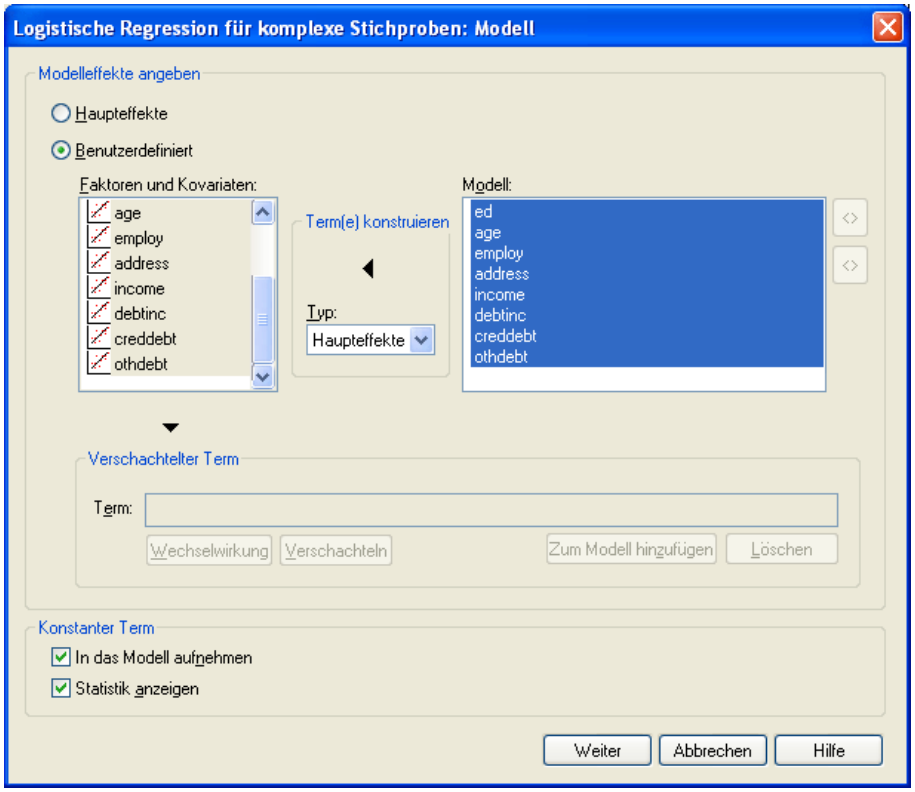
Abbildung 10-2
Dialogfeld "Logistische Regression: Referenzkategorie"



Bei der Prozedur “Logistische Regression für komplexe Stichproben” wird standardmäßig die höchstwertige Kategorie als Referenzkategorie verwendet. In diesem Dialogfeld können Sie den höchsten bzw. niedrigsten Wert oder auch eine benutzerdefinierte Kategorie als Referenzkategorie angeben.

Logistische Regression für komplexe Stichproben: Modell

Abbildung 10-3
Dialogfeld “Logistische Regression: Modell”



Modell-Effekte angeben. Standardmäßig erstellt das Verfahren ein Modell mit Haupteffekten unter Verwendung der im Hauptdialogfeld angegebenen Faktoren und Kovariaten. Alternativ können Sie ein benutzerdefiniertes Modell erstellen, das Wechselwirkungseffekte und verschachtelte Terme enthält.

Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung für alle ausgewählten Variablen erzeugt.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-fach. Hiermit werden alle möglichen 2-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-fach. Hiermit werden alle möglichen 3-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-fach. Hiermit werden alle möglichen 4-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-fach. Hiermit werden alle möglichen 5-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Verschachtelte Terme

In dieser Prozedur können Sie verschachtelte Terme für ein Modell konstruieren. Verschachtelte Terme sind nützlich, um den Effekt von Faktoren oder Kovariaten zu analysieren, deren Werte nicht mit den Stufen eines anderen Faktors interagieren. Eine Lebensmittelkette kann beispielsweise das Kaufverhalten ihrer Kunden in mehreren Filialen untersuchen. Da jeder Kunde nur eine dieser Filialen besucht, kann der Effekt *Kunde* als **verschachtelt innerhalb** des Effekts *Filiale* beschrieben werden.

Darüber hinaus können Sie Wechselwirkungseffekte, wie polynomiale Terme mit derselben Kovariaten, einschließen oder dem verschachtelten Term mehrere Verschachtelungsebenen hinzufügen.

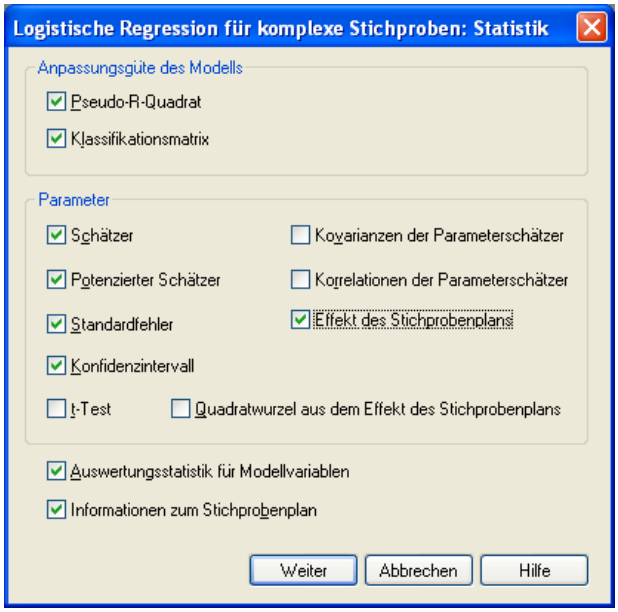
Einschränkungen. Für verschachtelte Terme gelten die folgenden Einschränkungen:

- Alle Faktoren innerhalb einer Wechselwirkung müssen eindeutig sein. Dementsprechend ist die Angabe von *A*A* unzulässig, wenn *A* ein Faktor ist.
- Alle Faktoren innerhalb eines verschachtelten Effekts müssen eindeutig sein. Dementsprechend ist die Angabe von *A(A)* unzulässig, wenn *A* ein Faktor ist.
- Effekte dürfen nicht in einer Kovariaten verschachtelt werden. Dementsprechend ist die Angabe von *A(X)* unzulässig, wenn *A* ein Faktor und *X* eine Kovariate ist.

Konstanter Term. Der konstante Term wird gewöhnlich in das Modell aufgenommen. Wenn anzunehmen ist, dass die Daten durch den Koordinatenursprung verlaufen, können Sie den konstanten Term ausschließen. Selbst wenn Sie den konstanten Term in das Modell aufnehmen, können Sie festlegen, dass die darauf bezogenen Statistiken unterdrückt werden sollen.

Logistische Regression für komplexe Stichproben: Statistik

Abbildung 10-4
Dialogfeld "Logistische Regression: Statistik"



Anpassungsgüte des Modells. Steuert die Anzeige der Statistik, in der die Gesamtleistung des Modells bewertet wird.

- **Pseudo-R-Quadrat.** Für die R^2 -Statistik aus der linearen Regression bieten die Modelle für die logistische Regression kein exaktes Gegenstück. Mit diesen Mehrfachmessungen werden stattdessen die Eigenschaften der R^2-Statistik nachgebildet.
- **Klassifikationsmatrix.** Zeigt die ausgewerteten Kreuzklassifikationen der beobachteten Kategorie nach der vom Modell vorhergesagten Kategorie für die abhängige Variable.

Parameter. In dieser Gruppe steuern Sie die Anzeige der Statistiken für die Modellparameter.

- **Schätzer.** Zeigt eine Schätzung der Koeffizienten.
- **Potenzierter Schätzer.** Zeigt die Basis des natürlichen Logarithmus, potenziert mit dem Schätzer der Koeffizienten. Der Schätzer bietet zwar ergiebige Eigenschaften für statische Tests; der potenzierte Schätzer oder exp(B) ist jedoch einfacher zu interpretieren.
- **Standardfehler.** Zeigt den Standardfehler für die einzelnen Koeffizientenschätzer.
- **Konfidenzintervall.** Zeigt ein Konfidenzintervall für die einzelnen Koeffizientenschätzer. Das Konfidenzniveau für das Intervall wird im Dialogfeld "Optionen" festgelegt.
- **T-Test.** Zeigt je einen *t*-Test für die einzelnen Koeffizientenschätzer. Die Nullhypothese der Tests liegt vor, wenn der Koeffizient den Wert 0 aufweist.
- **Kovarianzen der Parameterschätzer.** Zeigt eine Schätzung der Kovarianzmatrix für die Modellkoeffizienten.
- **Korrelationen der Parameterschätzer.** Zeigt eine Schätzung der Korrelationsmatrix für die Modellkoeffizienten.

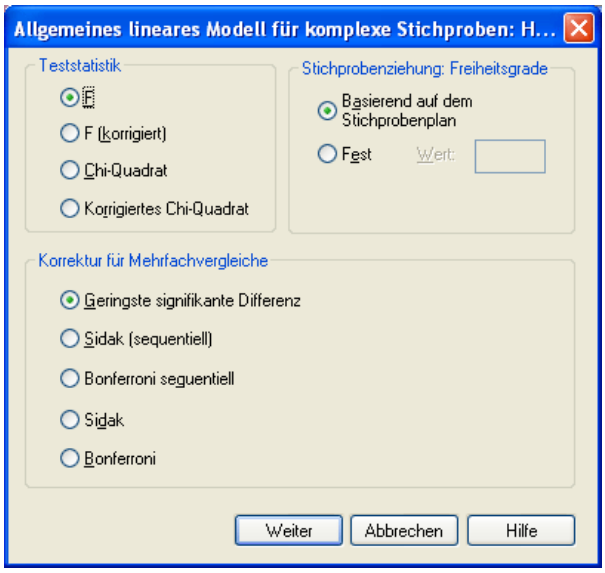
- **Effekt des Stichprobenplans.** Das Verhältnis der Varianz des Schätzers zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Es handelt sich hierbei um ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.
- **Quadratwurzel aus dem Effekt des Stichprobenplans.** Dies ist ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.

Auswertungsstatistik für Modellvariablen. Zeigt eine Zusammenfassung über die abhängige Variable, die Kovariaten und die Faktoren.

Informationen zum Stichprobenplan. Zeigt eine Zusammenfassung über die Stichprobe (mit ungewichteter Anzahl und Umfang der Grundgesamtheit).

Hypothesentests für komplexe Stichproben

Abbildung 10-5
Dialogfeld "Hypothesentests"



Teststatistik. In dieser Gruppe können Sie den Typ der Statistik zum Testen der Hypothesen festlegen. Die folgenden Optionen stehen zur Auswahl: *F*, *F* (korrigiert), "Chi-Quadrat" und "Korrigiertes Chi-Quadrat".

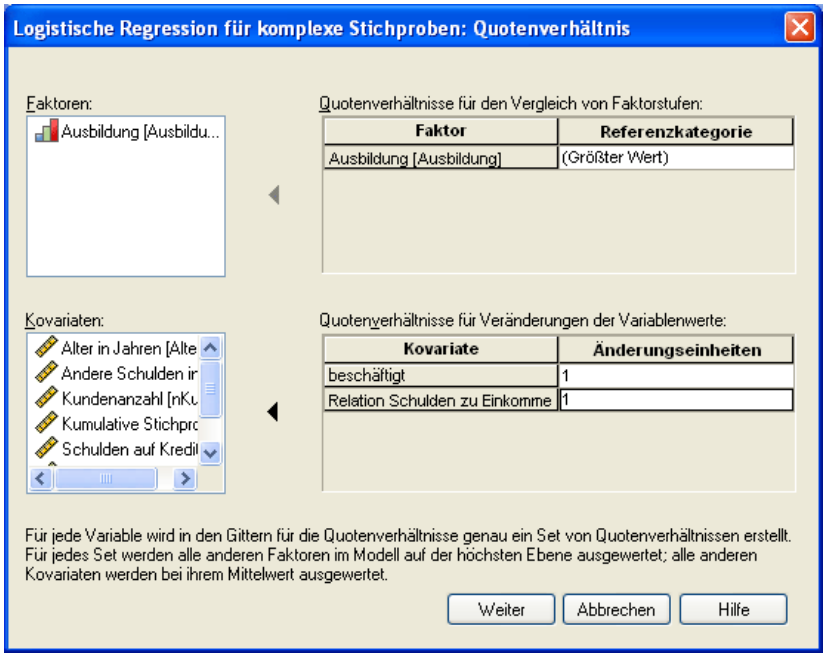
Stichprobenziehung: Freiheitsgrade. In dieser Gruppe steuern Sie die Freiheitsgrade im Stichprobenplan, mit denen die *p*-Werte für alle Teststatistiken berechnet werden. Dient der Stichprobenplan als Grundlage, ist dieser Wert die Differenz zwischen der Anzahl der primären Stichprobeneinheiten und der Anzahl der Schichten in der ersten Stufe der Stichproben. Alternativ können Sie benutzerdefinierte Freiheitsgrade festlegen; geben Sie hierzu eine positive Ganzzahl ein.

Korrektur für Mehrfachvergleiche. Bei der Durchführung von Hypothesentests mit mehreren Kontrasten kann das Gesamtsignifikanzniveau mithilfe der Signifikanzniveaus der eingeschlossenen Kontraste angepasst werden. In dieser Gruppe können Sie die Anpassungs-/Korrekturmethode auswählen.

- **Geringste signifikante Differenz.** Diese Methode steuert nicht die Gesamtwahrscheinlichkeit, dass Hypothesen abgelehnt werden, bei denen einige lineare Kontraste von den Werten einer Nullhypothese abweichen.
- **Sidak (sequenziell).** Hierbei handelt es sich um ein sequenzielles schrittweises Sidak-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen anbelangt, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.
- **Bonferroni sequenziell.** Hierbei handelt es sich um ein sequenzielles schrittweises Bonferroni-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen anbelangt, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.
- **Sidak.** Dieses Verfahren liefert engere Grenzen als der Bonferroni-Ansatz.
- **Bonferroni.** Dieses Verfahren passt das empirische Signifikanzniveau der Tatsache an, dass mehrere Kontraste getestet werden.

Logistische Regression für komplexe Stichproben: Quotenverhältnis

Abbildung 10-6
Dialogfeld "Logistische Regression: Quotenverhältnisse"



Im Dialogfeld "Quotenverhältnisse" rufen Sie die vom Modell geschätzten Quotenverhältnisse für bestimmte Faktoren und Kovariaten ab. Für jede Kategorie der abhängigen Variable (mit Ausnahme der Referenzkategorie) wird je ein separater Satz von Quotenverhältnissen berechnet.

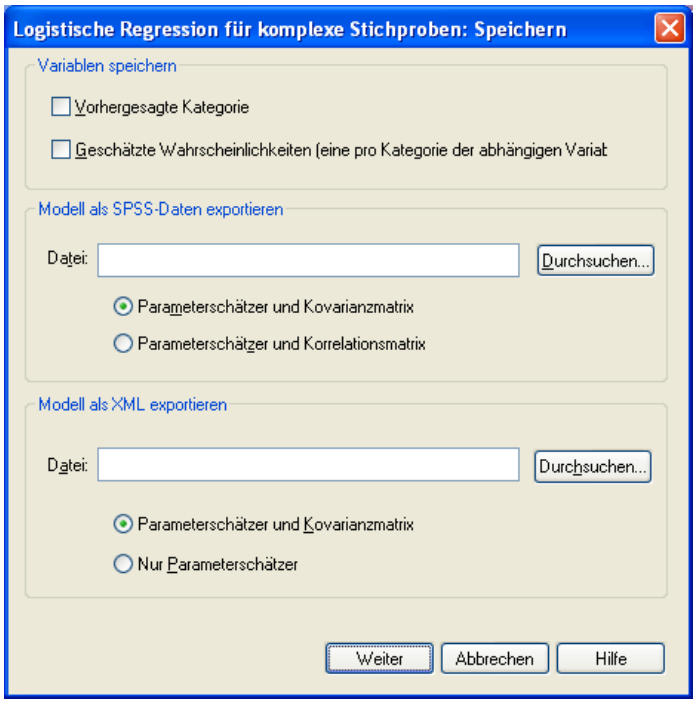
Faktoren. Für jeden ausgewählten Faktor wird das Verhältnis der Quoten in jeder Kategorie des Faktors zu den Quoten in der angegebenen Referenzkategorie angezeigt.

Kovariaten. Für jede ausgewählte Kovariate wird das Verhältnis der Quoten für den Mittelwert der Kovariate zzgl. der angegebenen Änderungseinheiten zu den Quoten für den Mittelwert angezeigt.

Beim Berechnen der Quotenverhältnisse für einen Faktor oder eine Kovariate werden alle anderen Faktoren auf die jeweils höchste Ebene fixiert, alle anderen Kovariaten dagegen auf den Mittelwert. Liegen Wechselwirkungen eines Faktors oder einer Kovariate mit anderen Einflussgrößen im Modell vor, sind die Quotenverhältnisse nicht nur von den Änderungen bei der angegebenen Variable abhängig, sondern auch von den Werten der anderen Variablen, mit denen die Wechselwirkungen bestehen. Wenn eine angegebene Kovariate im Modell Wechselwirkungen mit sich selbst aufweist (z. B. *alter*alter*), sind die Quotenverhältnisse sowohl von den Änderungen bei der Kovariate abhängig als auch vom Wert der Kovariate.

Logistische Regression für komplexe Stichproben: Speichern

Abbildung 10-7
Dialogfeld "Logistische Regression: Speichern"



Variablen speichern. In dieser Gruppe speichern Sie die vom Modell vorhergesagte Kategorie und die vorhergesagten Wahrscheinlichkeiten als neue Variablen in der Arbeitsdatei.

Modell als SPSS-Daten exportieren. Schreibt ein SPSS-Daten-Set, das die Parameter-Korrelations- oder -Kovarianzmatrix enthält (mit Parameterschätzern, Standardfehlern, Signifikanzwerten und Freiheitsgraden). Die Reihenfolge der Variablen in der Matrixdatei lautet wie folgt:

- **rowtype_.** Nimmt folgende Werte (und Wertelabel) an: COV (Kovarianzen), CORR (Korrelationen), EST (Parameterschätzer), SE (Standardfehler), SIG (Signifikanzniveaus) und DF (Freiheitsgrade des Stichprobenplans). Es gibt einen separaten Fall mit dem Zeilentyp COV (bzw. CORR) für jeden Modellparameter sowie einen separaten Fall für jeden der anderen Zeilentypen.

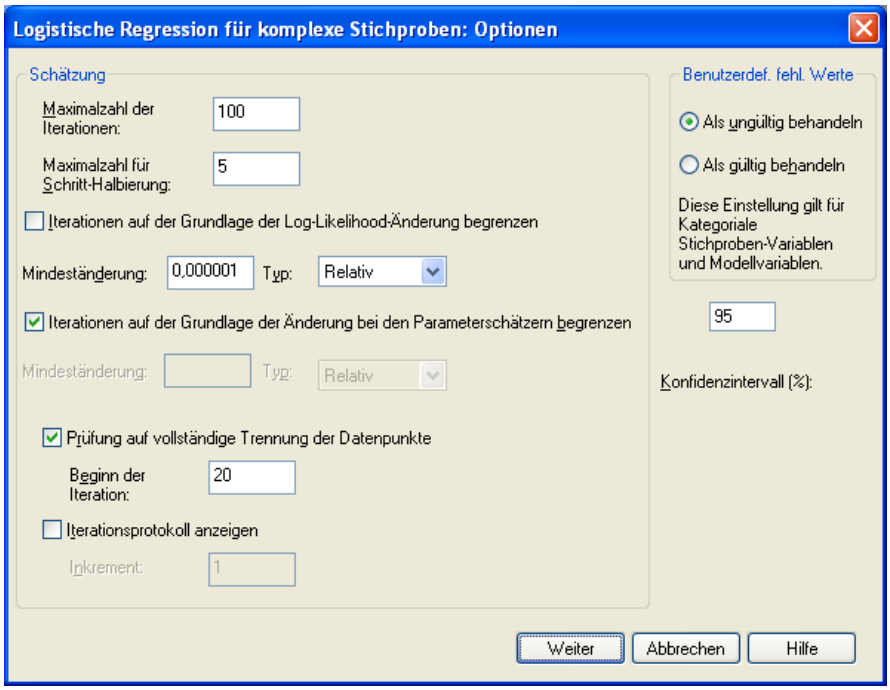
- **varname_.** Nimmt für die Zeilentypen COV bzw. CORR die Werte P1, P2, ... an, was einer geordneten Liste aller geschätzten Modellparameter entspricht, mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Für alle anderen Zeilentypen sind die Zellen leer.
- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Die Werte hängen jeweils vom Zeilentyp ab. Bei redundanten Parametern sind alle Kovarianzen auf 0 gesetzt, die Korrelationen sind auf den systemdefiniert fehlenden Wert gesetzt, alle Parameterschätzer sind auf 0 gesetzt und alle Standardfehler, Signifikanzniveaus und die Freiheitsgrade der Residuen sind auf den systemdefiniert fehlenden Wert gesetzt.

Anmerkung: Diese Datei ist nicht unmittelbar für weitere Analysen in anderen Prozeduren verwendbar, bei denen eine Matrixdatei eingelesen wird, es sei denn, diese Prozeduren akzeptieren alle hier exportierten Zeilentypen.

Modell als XML exportieren. Speichert die Parameterschätzungen und ggf. die Parameter-Kovarianzmatrix (falls ausgewählt) im XML-Format (PMML). SmartScore und SPSS Server (gesondertes Produkt) können anhand dieser Modelldatei die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.

Logistische Regression für komplexe Stichproben: Optionen

Abbildung 10-8
Dialogfeld "Logistische Regression: Optionen"



Schätzung. In dieser Gruppe steuern Sie verschiedene Kriterien für die Schätzung im Modell.

- **Maximalzahl der Iterationen.** Dies ist die maximale Anzahl der Iterationen, die im Algorithmus vorgenommen werden. Geben Sie eine nichtnegative Ganzzahl an.

- **Maximalzahl für Schritt-Halbierung.** Bei jeder Iteration wird die Schrittgröße um den Faktor 0,5 reduziert, bis die Log-Likelihood ansteigt oder die Maximalzahl für die Schritt-Halbierung erreicht ist. Geben Sie eine positive Ganzzahl ein.
- **Iterationen auf der Grundlage der Änderung bei den Parameterschätzern begrenzen.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzern unter dem angegeben (nicht negativen) Wert liegt.
- **Iterationen auf der Grundlage der Log-Likelihood-Änderung begrenzen.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei der Log-Likelihood-Funktion unter dem angegeben (nicht negativen) Wert liegt.
- **Prüfung auf vollständige Trennung der Datenpunkte.** Mit dieser Option lassen Sie Tests durch den Algorithmus durchführen, mit denen sichergestellt wird, dass die Parameterschätzer eindeutige Werte aufweisen. Eine Trennung wird vorgenommen, sobald ein Modell erzeugt werden kann, in dem alle Fälle fehlerfrei klassifiziert werden.
- **Iterationsprotokoll anzeigen.** Die Parameterschätzer und die Statistik werden alle *n* Iterationen angezeigt, beginnend mit der 0. Iteration (den ursprünglichen Schätzungen). Wenn Sie das Iterationsprotokoll drucken, wird die letzte Iteration stets unabhängig vom Wert für *n* ausgegeben.

Benutzerdefinierte fehlende Werte. Alle Stichproben-Variablen sowie die abhängige Variable und ggf. alle Kovariaten müssen gültige Daten enthalten. Fälle, bei denen ungültige Daten für diese Variablen vorliegen, werden aus der Analyse gelöscht. Mit diesen Steuerungen legen Sie fest, ob benutzerdefiniert fehlende Werte bei den Schicht-, Cluster-, Teilgesamtheits- und Faktorvariablen als gültige Werte behandelt werden sollen.

Konfidenzintervall. Dies ist die Konfidenzintervall-Ebene für Koeffizientenschätzungen, potenzierte Koeffizientenschätzungen und Quotenverhältnisse. Geben Sie einen Wert größer oder gleich 50 und kleiner als 100 ein.

Zusätzliche Funktionen beim Befehl CSLOGISTIC

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Mit dem Unterbefehl `CUSTOM` können Sie benutzerdefinierte Tests auf Effekte im Vergleich zu linearen Kombinationen von Effekten oder einem Wert vornehmen.
- Mit dem Unterbefehl `ODDSRATIOS` können Sie Werte für andere Modellvariablen festlegen, wenn Sie die Quotenverhältnisse für Faktoren und Kovariaten berechnen.
- Mit dem Unterbefehl `CRITERIA` können Sie einen Toleranzwert für die Prüfung auf Singularität festlegen.
- Mit dem Unterbefehl `SAVE` können Sie benutzerdefinierte Namen für gespeicherte Variablen angeben.
- Mit dem Unterbefehl `PRINT` können Sie eine Tabelle mit allgemeinen schätzbaren Funktionen anlegen.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
11

Ordinale Regression für komplexe Stichproben

Die Prozedur "Ordinale Regression für komplexe Stichproben" besteht aus einer Regressionsanalyse einer binären oder ordinalen abhängigen Variablen für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Optional können Sie auch Analysen für eine Teilgesamtheit vornehmen.

Beispiel. Abgeordnete, die in Erwägung ziehen, einen Gesetzesentwurf einzubringen, sind daran interessiert zu ermitteln, ob dieser Gesetzesantrag öffentlich unterstützt wird und in welchem Bezug die Unterstützung für den Antrag zur demografischen Struktur der Wähler steht. Die Meinungsforscher verwenden für die Erstellung und Durchführung der entsprechenden Umfragen einen komplexen Stichprobenplan. Mit der ordinalen Regression für komplexe Stichproben können Sie ein Modell für die Stärke der Unterstützung für den Gesetzesentwurf auf der Grundlage der demografischen Struktur der Wähler anpassen.

Daten. Die abhängige Variable ist ordinal. Faktoren sind kategorial. Kovariaten sind quantitative Variablen, die mit der abhängigen Variablen in Beziehung stehen. Bei den Variablen für die Teilgesamtheiten kann es sich um String-Variablen oder numerische Variablen handeln, sie sollten jedoch kategorial sein.

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Plan, deren Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld "Komplexe Stichproben: Plan" ausgewählt wurde.

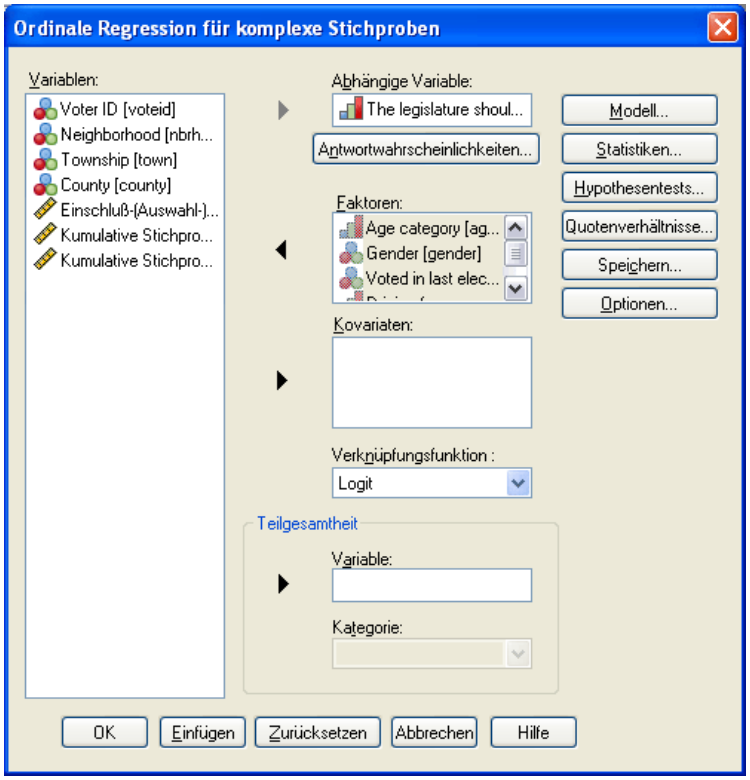
Ermitteln der ordinalen Regression für komplexe Stichproben

Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Komplexe Stichproben
    Ordinale Regression...

- Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.
- Klicken Sie auf Weiter.

Abbildung 11-1
Dialogfeld "Ordinale Regression"



▶ Wählen Sie eine abhängige Variable aus.

Die folgenden Optionen sind verfügbar:

- Wählen Sie die entsprechenden Variablen für Faktoren und Kovariaten gemäß den vorliegenden Daten aus.
- Legen Sie eine Variable fest, um so eine Teilgesamtheit zu definieren. Die Analyse wird ausschließlich für die ausgewählte Kategorie der Teilgesamtheit-Variable vorgenommen, die Varianzen jedoch werden dennoch ordnungsgemäß auf der Grundlage des gesamten Daten-Sets geschätzt.
- Wählen Sie eine Link-Funktion.

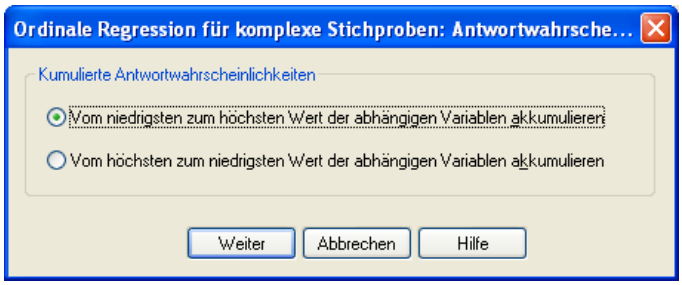
Link-Funktion. Die Link-Funktion ist eine Transformation der kumulativen Wahrscheinlichkeiten, die eine Schätzung des Modells ermöglicht. Es stehen fünf Link-Funktionen zur Verfügung, die in der folgenden Tabelle zusammengefasst sind.

Funktion	**Form**	**Typische Anwendung**
Logit	$\log(\xi / (1-\xi))$	Gleichmäßig verteilte Kategorien
Log-Log komplementär	$\log(-\log(1-\xi))$	Höhere Kategorien wahrscheinlicher
Log-Log negativ	$-\log(-\log(\xi))$	Niedrigere Kategorien wahrscheinlicher

Funktion	Form	Typische Anwendung
Probit	$\Phi^{-1}(\xi)$	Latente Variable ist normalverteilt
Cauchit (Inverse von Cauchy)	$\tan(\pi(\xi-0,5))$	Latente Variable weist viele Extremwerte auf

Ordinale Regression für komplexe Stichproben: Antwortwahrscheinlichkeiten

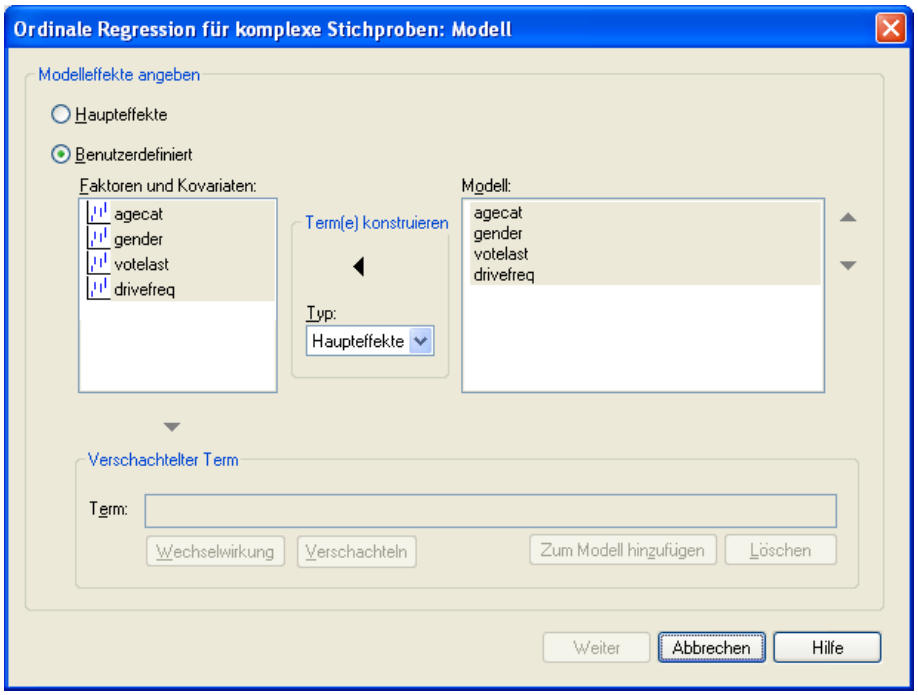
Abbildung 11-2
Dialogfeld "Ordinale Regression: Antwortwahrscheinlichkeiten"


Ordinale Regression für komplexe Stichproben: Antwortwahrsche...
Kumulierte Antwortwahrscheinlichkeiten
Vom niedrigsten zum höchsten Wert der abhängigen Variablen akkumulieren
Vom höchsten zum niedrigsten Wert der abhängigen Variablen akkumulieren
Weiter
Abbrechen
Hilfe

Im Dialogfeld "Antwortwahrscheinlichkeiten" können Sie angeben, ob die kumulative Wahrscheinlichkeit einer Antwort (d. h. die Wahrscheinlichkeit der Zugehörigkeit bis zu einschließlich einer bestimmten Kategorie der abhängigen Variablen) mit steigenden bzw. sinkenden Werten der abhängigen Variablen steigt.

Ordinale Regression für komplexe Stichproben: Modell

Abbildung 11-3
Dialogfeld "Ordinale Regression: Modell"



Modell-Effekte angeben. Standardmäßig erstellt das Verfahren ein Modell mit Haupteffekten unter Verwendung der im Hauptdialogfeld angegebenen Faktoren und Kovariaten. Alternativ können Sie ein benutzerdefiniertes Modell erstellen, das Wechselwirkungseffekte und verschachtelte Terme enthält.

Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung für alle ausgewählten Variablen erzeugt.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-fach. Hiermit werden alle möglichen 2-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-fach. Hiermit werden alle möglichen 3-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-fach. Hiermit werden alle möglichen 4-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-fach. Hiermit werden alle möglichen 5-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Verschachtelte Terme

In dieser Prozedur können Sie verschachtelte Terme für ein Modell konstruieren. Verschachtelte Terme sind nützlich, um den Effekt von Faktoren oder Kovariaten zu analysieren, deren Werte nicht mit den Stufen eines anderen Faktors interagieren. Eine Lebensmittelkette kann beispielsweise das Kaufverhalten ihrer Kunden in mehreren Filialen untersuchen. Da jeder Kunde nur eine dieser Filialen besucht, kann der Effekt *Kunde* als **verschachtelt innerhalb** des Effekts *Filiale* beschrieben werden.

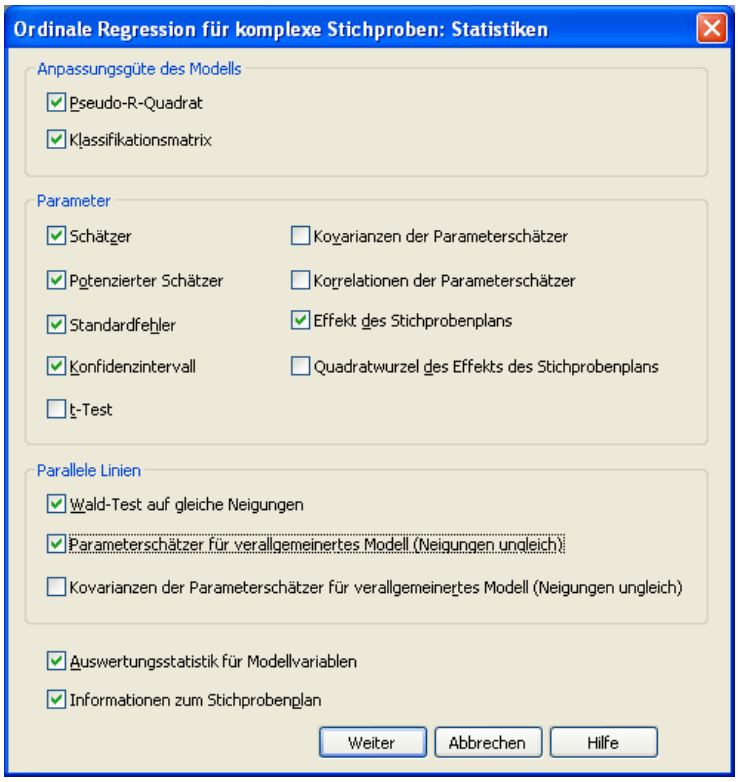
Darüber hinaus können Sie Wechselwirkungseffekte, wie polynomiale Terme mit derselben Kovariaten, einschließen oder dem verschachtelten Term mehrere Verschachtelungsebenen hinzufügen.

Einschränkungen. Für verschachtelte Terme gelten die folgenden Einschränkungen:

- Alle Faktoren innerhalb einer Wechselwirkung müssen eindeutig sein. Dementsprechend ist die Angabe von *A*A* unzulässig, wenn *A* ein Faktor ist.
- Alle Faktoren innerhalb eines verschachtelten Effekts müssen eindeutig sein. Dementsprechend ist die Angabe von *A(A)* unzulässig, wenn *A* ein Faktor ist.
- Effekte dürfen nicht in einer Kovariaten verschachtelt werden. Dementsprechend ist die Angabe von *A(X)* unzulässig, wenn *A* ein Faktor und *X* eine Kovariate ist.

Ordinale Regression für komplexe Stichproben: Statistik

Abbildung 11-4
Dialogfeld "Ordinale Regression: Statistik"



Anpassungsgüte des Modells. Steuert die Anzeige der Statistik, in der die Gesamtleistung des Modells bewertet wird.

- **Pseudo-R-Quadrat.** Für die R^2-Statistik aus der linearen Regression bieten die Modelle für die ordinale Regression kein exaktes Gegenstück. Mit diesen Mehrfachmessungen werden stattdessen die Eigenschaften der R^2-Statistik nachgebildet.
- **Klassifikationsmatrix.** Zeigt die ausgewerteten Kreuzklassifikationen der beobachteten Kategorie nach der vom Modell vorhergesagten Kategorie für die abhängige Variable.

Parameter. In dieser Gruppe steuern Sie die Anzeige der Statistiken für die Modellparameter.

- **Schätzer.** Zeigt eine Schätzung der Koeffizienten.
- **Potenzierter Schätzer.** Zeigt die Basis des natürlichen Logarithmus, potenziert mit dem Schätzer der Koeffizienten. Der Schätzer bietet zwar ergiebige Eigenschaften für statische Tests; der potenzierte Schätzer oder exp(B) ist jedoch einfacher zu interpretieren.
- **Standardfehler.** Zeigt den Standardfehler für die einzelnen Koeffizientenschätzer.
- **Konfidenzintervall.** Zeigt ein Konfidenzintervall für die einzelnen Koeffizientenschätzer. Das Konfidenzniveau für das Intervall wird im Dialogfeld “Optionen” festgelegt.
- **T-Test.** Zeigt je einen *t*-Test für die einzelnen Koeffizientenschätzer. Die Nullhypothese der Tests liegt vor, wenn der Koeffizient den Wert 0 aufweist.
- **Kovarianzen der Parameterschätzer.** Zeigt eine Schätzung der Kovarianzmatrix für die Modellkoeffizienten.
- **Korrelationen der Parameterschätzer.** Zeigt eine Schätzung der Korrelationsmatrix für die Modellkoeffizienten.
- **Effekt des Stichprobenplans.** Das Verhältnis der Varianz des Schätzers zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Es handelt sich hierbei um ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.
- **Quadratwurzel aus dem Effekt des Stichprobenplans.** Dies ist ein Maß für den Effekt der Angabe eines komplexen Plans (ausgedrückt in Einheiten, die denen des Standardfehlers vergleichbar sind). Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.

Parallele Linien. In dieser Gruppe können Sie Statistiken anfordern, die einem Modell mit nichtparallelen Linien zugeordnet sind. Dabei wird eine separate Regressionslinie für jede Antwortkategorie (außer der letzten) angepasst.

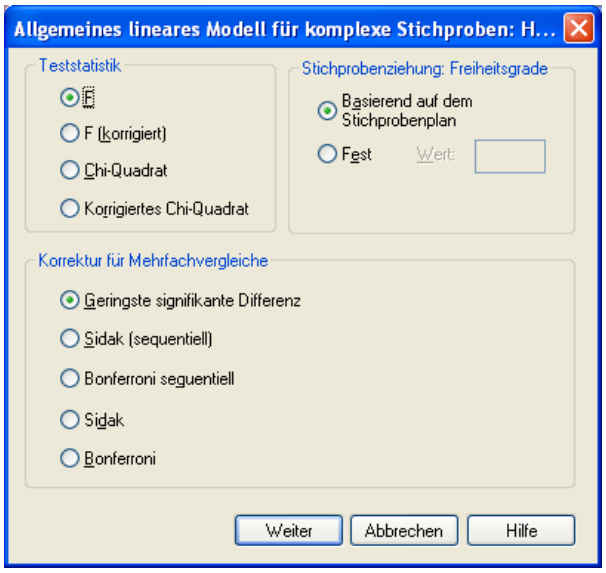
- **Wald-Test.** Erstellt einen Test für die Nullhypothese, dass die Regressionsparameter für alle kumulativen Antworten gleich sind. Das Modell mit nichtparallelen Linien wird geschätzt und der Wald-Test auf gleiche Parameter wird angewendet.
- **Parameterschätzer.** Zeigt Schätzwerte für die Koeffizienten und Standardfehler des Modells mit nichtparallelen Linien an.
- **Kovarianzen der Parameterschätzer.** Zeigt eine Schätzung der Kovarianzmatrix für die Koeffizienten des Modells mit nichtparallelen Linien an.

Auswertungsstatistik für Modellvariablen. Zeigt eine Zusammenfassung über die abhängige Variable, die Kovariaten und die Faktoren.

Informationen zum Stichprobenplan. Zeigt eine Zusammenfassung über die Stichprobe (mit ungewichteter Anzahl und Umfang der Grundgesamtheit).

Hypothesentests für komplexe Stichproben

Abbildung 11-5
Dialogfeld "Hypothesentests"



Teststatistik. In dieser Gruppe können Sie den Typ der Statistik zum Testen der Hypothesen festlegen. Die folgenden Optionen stehen zur Auswahl: *F*, *F* (korrigiert), "Chi-Quadrat" und "Korrigiertes Chi-Quadrat".

Stichprobenziehung: Freiheitsgrade. In dieser Gruppe steuern Sie die Freiheitsgrade im Stichprobenplan, mit denen die *p*-Werte für alle Teststatistiken berechnet werden. Dient der Stichprobenplan als Grundlage, ist dieser Wert die Differenz zwischen der Anzahl der primären Stichprobeneinheiten und der Anzahl der Schichten in der ersten Stufe der Stichproben. Alternativ können Sie benutzerdefinierte Freiheitsgrade festlegen; geben Sie hierzu eine positive Ganzzahl ein.

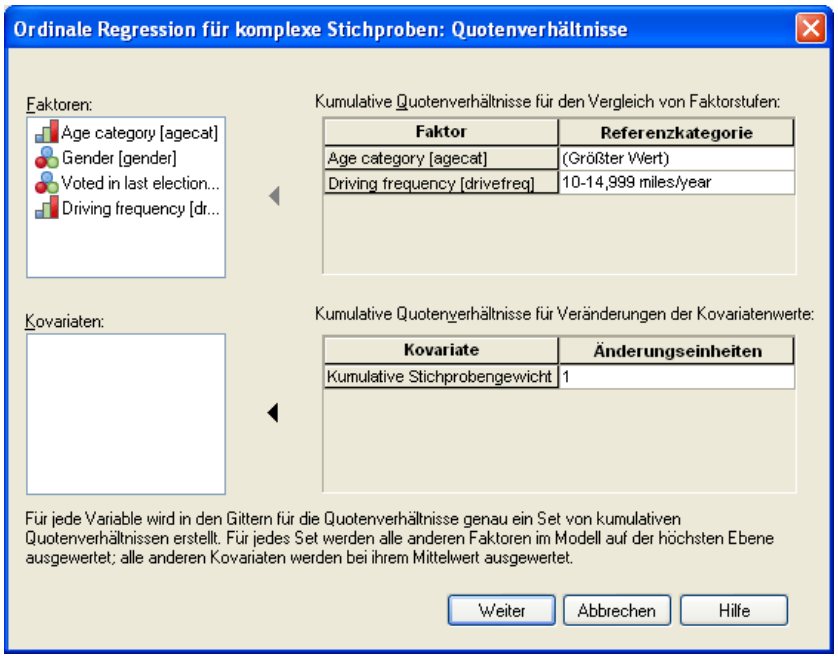
Korrektur für Mehrfachvergleiche. Bei der Durchführung von Hypothesentests mit mehreren Kontrasten kann das Gesamtsignifikanzniveau mithilfe der Signifikanzniveaus der eingeschlossenen Kontraste angepasst werden. In dieser Gruppe können Sie die Anpassungs-/Korrekturmethode auswählen.

- **Geringste signifikante Differenz.** Diese Methode steuert nicht die Gesamtwahrscheinlichkeit, dass Hypothesen abgelehnt werden, bei denen einige lineare Kontraste von den Werten einer Nullhypothese abweichen.
- **Sidak (sequenziell).** Hierbei handelt es sich um ein sequenzielles schrittweises Sidak-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen anbelangt, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.
- **Bonferroni sequenziell.** Hierbei handelt es sich um ein sequenzielles schrittweises Bonferroni-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen anbelangt, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.

- **Sidak.** Dieses Verfahren liefert engere Grenzen als der Bonferroni-Ansatz.
- **Bonferroni.** Dieses Verfahren passt das empirische Signifikanzniveau der Tatsache an, dass mehrere Kontraste getestet werden.

Ordinale Regression für komplexe Stichproben: Quotenverhältnisse

Abbildung 11-6
Dialogfeld "Ordinale Regression: Quotenverhältnisse"


Ordinale Regression für komplexe Stichproben: Quotenverhältnisse
Faktoren:
Age category [agecat]
Gender [gender]
Voted in last election...
Driving frequency [dr...
Kumulative Quotenverhältnisse für den Vergleich von Faktorstufen:
Faktor
Referenzkategorie
Age category [agecat]
(Größter Wert)
Driving frequency [drivefreq]
10-14,999 miles/year
Kovariaten:
Kumulative Quotenverhältnisse für Veränderungen der Kovariatenwerte:
Kovariate
Änderungseinheiten
Kumulative Stichprobengewicht
1
Für jede Variable wird in den Gittern für die Quotenverhältnisse genau ein Set von kumulativen Quotenverhältnissen erstellt. Für jedes Set werden alle anderen Faktoren im Modell auf der höchsten Ebene ausgewertet; alle anderen Kovariaten werden bei ihrem Mittelwert ausgewertet.
Weiter
Abbrechen
Hilfe

Im Dialogfeld "Quotenverhältnisse" rufen Sie die vom Modell geschätzten kumulativen Quotenverhältnisse für bestimmte Faktoren und Kovariaten ab. Diese Funktion ist nur bei Modellen verfügbar, die die Link-Funktion "Logit" verwenden. Für alle Kategorien der abhängigen Variablen mit Ausnahme der letzten wird ein einziges kumulatives Quotenverhältnis berechnet; das proportionale Odds-Modell postuliert, dass alle gleich sind.

Faktoren. Für jeden ausgewählten Faktor wird das Verhältnis der kumulativen Quoten (Odds) in jeder Kategorie des Faktors zu den Quoten in der angegebenen Referenzkategorie angezeigt.

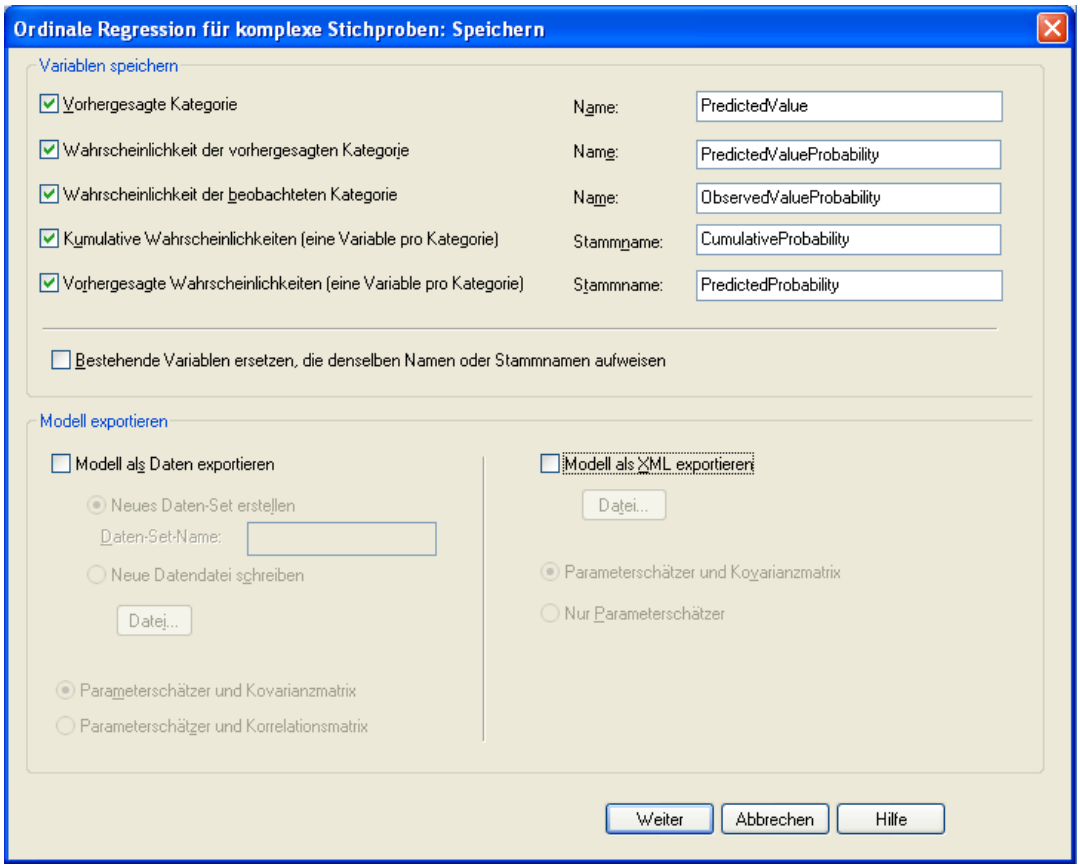
Kovariaten. Für jede ausgewählte Kovariate wird das Verhältnis der kumulativen Quoten (Odds) für den Mittelwert der Kovariaten zzgl. der angegebenen Änderungseinheiten zu den Quoten für den Mittelwert angezeigt.

Beim Berechnen der Quotenverhältnisse für einen Faktor oder eine Kovariate werden alle anderen Faktoren auf die jeweils höchste Ebene fixiert, alle anderen Kovariaten dagegen auf den Mittelwert. Liegen Wechselwirkungen eines Faktors oder einer Kovariate mit anderen Einflussgrößen im Modell vor, sind die Quotenverhältnisse nicht nur von den Änderungen bei der angegebenen Variable abhängig, sondern auch von den Werten der anderen Variablen, mit denen die Wechselwirkungen bestehen. Wenn eine angegebene Kovariate im Modell

Wechselwirkungen mit sich selbst aufweist (z. B. *alter*alter*), sind die Quotenverhältnisse sowohl von den Änderungen bei der Kovariate abhängig als auch vom Wert der Kovariate.

Ordinale Regression für komplexe Stichproben: Speichern

Abbildung 11-7
Dialogfeld "Ordinale Regression: Speichern"



Variablen speichern. In dieser Gruppe speichern Sie die vom Modell vorhergesagte Kategorie, die Wahrscheinlichkeit der vorhergesagten Kategorie, die Wahrscheinlichkeit der beobachteten Kategorie und die vorhergesagten Wahrscheinlichkeiten als neue Variablen in der Arbeitsdatei.

Modell als SPSS-Daten exportieren. Schreibt ein SPSS-Daten-Set, das die Parameter-Korrelations- oder -Kovarianzmatrix enthält (mit Parameterschätzern, Standardfehlern, Signifikanzwerten und Freiheitsgraden). Die Reihenfolge der Variablen in der Matrixdatei lautet wie folgt:

- **rowtype_.** Nimmt folgende Werte (und Wertelabel) an: COV (Kovarianzen), CORR (Korrelationen), EST (Parameterschätzer), SE (Standardfehler), SIG (Signifikanzniveaus) und DF (Freiheitsgrade des Stichprobenplans). Es gibt einen separaten Fall mit dem Zeilentyp COV (bzw. CORR) für jeden Modellparameter sowie einen separaten Fall für jeden der anderen Zeilentypen.

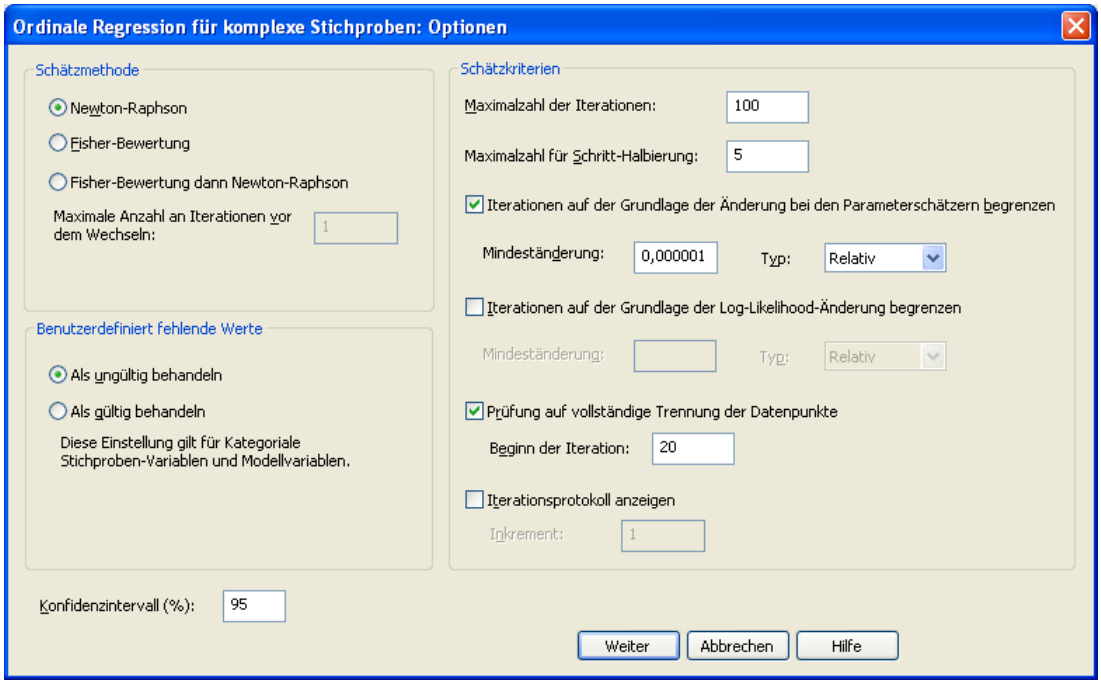
- **varname_.** Nimmt für die Zeilentypen COV bzw. CORR die Werte P1, P2, ... an, was einer geordneten Liste aller geschätzten Modellparameter entspricht, mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Für alle anderen Zeilentypen sind die Zellen leer.
- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Die Werte hängen jeweils vom Zeilentyp ab. Bei redundanten Parametern sind alle Kovarianzen auf 0 gesetzt, die Korrelationen sind auf den systemdefiniert fehlenden Wert gesetzt, alle Parameterschätzer sind auf 0 gesetzt und alle Standardfehler, Signifikanzniveaus und die Freiheitsgrade der Residuen sind auf den systemdefiniert fehlenden Wert gesetzt.

Anmerkung: Diese Datei ist nicht unmittelbar für weitere Analysen in anderen Prozeduren verwendbar, bei denen eine Matrixdatei eingelesen wird, es sei denn, diese Prozeduren akzeptieren alle hier exportierten Zeilentypen.

Modell als XML exportieren. Speichert die Parameterschätzungen und ggf. die Parameter-Kovarianzmatrix (falls ausgewählt) im XML-Format (PMML). SmartScore und SPSS Server (gesondertes Produkt) können anhand dieser Modelldatei die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.

Ordinale Regression für komplexe Stichproben: Optionen

Abbildung 11-8
Dialogfeld "Ordinale Regression: Optionen"



Schätzmethode. Sie können eine Methode für die Parameterschätzung auswählen. Sie haben die Wahl zwischen "Newton-Raphson", "Fisher-Bewertung" und einer Hybridmethode, bei der zuerst Iterationen der Fisher-Bewertung durchgeführt werden und dann zur Methode "Newton-Raphson"

gewechselt wird. Wenn während der Phase “Fisher-Bewertung” der Hybridmethode Konvergenz erreicht wird, bevor die maximale Anzahl an Fisher-Iterationen erreicht wurde, fährt der Algorithmus mit der Newton-Raphson-Methode fort.

Schätzung. In dieser Gruppe steuern Sie verschiedene Kriterien für die Schätzung im Modell.

- **Maximalzahl der Iterationen.** Dies ist die maximale Anzahl der Iterationen, die im Algorithmus vorgenommen werden. Geben Sie eine nichtnegative Ganzzahl an.
- **Maximalzahl für Schritt-Halbierung.** Bei jeder Iteration wird die Schrittgröße um den Faktor 0,5 reduziert, bis die Log-Likelihood ansteigt oder die Maximalzahl für die Schritt-Halbierung erreicht ist. Geben Sie eine positive Ganzzahl ein.
- **Iterationen auf der Grundlage der Änderung bei den Parameterschätzern begrenzen.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzern unter dem angegeben (nicht negativen) Wert liegt.
- **Iterationen auf der Grundlage der Log-Likelihood-Änderung begrenzen.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei der Log-Likelihood-Funktion unter dem angegeben (nicht negativen) Wert liegt.
- **Prüfung auf vollständige Trennung der Datenpunkte.** Mit dieser Option lassen Sie Tests durch den Algorithmus durchführen, mit denen sichergestellt wird, dass die Parameterschätzer eindeutige Werte aufweisen. Eine Trennung wird vorgenommen, sobald ein Modell erzeugt werden kann, in dem alle Fälle fehlerfrei klassifiziert werden.
- **Iterationsprotokoll anzeigen.** Die Parameterschätzer und die Statistik werden alle *n* Iterationen angezeigt, beginnend mit der 0. Iteration (den ursprünglichen Schätzungen). Wenn Sie das Iterationsprotokoll drucken, wird die letzte Iteration stets unabhängig vom Wert für *n* ausgegeben.

Benutzerdefinierte fehlende Werte. Metrische Stichproben-Variablen sowie die abhängige Variable und ggf. alle Kovariaten müssen gültige Daten enthalten. Fälle, bei denen ungültige Daten für diese Variablen vorliegen, werden aus der Analyse gelöscht. Mit diesen Steuerungen legen Sie fest, ob benutzerdefiniert fehlende Werte bei den Schicht-, Cluster-, Teilgesamtheits- und Faktorvariablen als gültige Werte behandelt werden sollen.

Konfidenzintervall. Dies ist die Konfidenzintervall-Ebene für Koeffizientenschätzungen, potenzierte Koeffizientenschätzungen und Quotenverhältnisse. Geben Sie einen Wert größer oder gleich 50 und kleiner als 100 ein.

Zusätzliche Funktionen beim Befehl CSORDINAL

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Mit dem Unterbefehl `CUSTOM` können Sie benutzerdefinierte Tests auf Effekte im Vergleich zu linearen Kombinationen von Effekten oder einem Wert vornehmen.
- Mit dem Unterbefehl `ODDSRATIOS` können Sie andere Werte als die Mittelwerte für andere Modellvariablen festlegen, wenn Sie die kumulativen Quotenverhältnisse für Faktoren und Kovariaten berechnen.
- Mit dem Unterbefehl `ODDSRATIOS` können Sie Werte ohne Label als benutzerdefinierte Referenzkategorien für Faktoren verwenden, wenn Quotenverhältnisse angefordert werden.

- Mit dem Unterbefehl `CRITERIA` können Sie einen Toleranzwert für die Prüfung auf Singularität festlegen.
- Mit dem Unterbefehl `PRINT` können Sie eine Tabelle mit allgemeinen schätzbaren Funktionen anlegen.
- Mit dem Unterbefehl `SAVE` können Sie mehr als 25 Wahrscheinlichkeitsvariablen speichern.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
12

Cox-Regression für komplexe Stichproben

Die Prozedur "Cox-Regression für komplexe Stichproben" besteht aus einer Überlebensanalyse für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Optional können Sie auch Analysen für eine Teilgesamtheit vornehmen.

Beispiele. Eine Strafverfolgungsbehörde ist hinsichtlich der Rückfallraten in ihrem Zuständigkeitsbereich unsicher. Einer der Messwerte der Rückfallrate ist die Zeit bis zur zweiten Festnahme von Straftätern. Die Behörde möchte die Zeit bis zur erneuten Festnahme mithilfe der Cox-Regression modellieren, ist jedoch besorgt, dass die proportionale Hazard-Annahme für die einzelnen Alterskategorien ungültig ist.

Medizinforscher untersuchen die Überlebenszeiten von Patienten nach einem Rehabilitationsprogramm wegen eines ischämischen Schlaganfalls. Es ist Potenzial für mehrere Fälle pro Subjekt vorhanden, da sich Anamnesen ändern, wenn das Auftreten von signifikanten Ereignissen mit nicht tödlichem Ausgang aufgezeichnet und die Zeiten dieser Ereignisse erfasst werden. Die Stichprobe ist auch auf der linken Seite in dem Sinne abgeschnitten, dass die beobachteten Überlebenszeiten durch die Rehabilitationslänge "erhöht" sind, da das Risiko zwar zum Zeitpunkt des ischämischen Schlaganfalls beginnt, in der Stichprobe jedoch nur Patienten aufgeführt sind, die auch noch nach dem Rehabilitationsprogramm am Leben sind.

Überlebenszeit. Die Prozedur wendet die Cox-Regression auf die Analyse von Überlebenszeiten an — d. h. also, die Dauer vor dem Auftreten eines Ereignisses. Es gibt zwei Arten zum Angeben der Überlebenszeit, je nach der Startzeit des Intervalls:

- **Zeit = 0.** Sie verfügen im Allgemeinen über alle Informationen zu Beginn des Intervalls für jedes Subjekt sowie einfach über eine Variable mit Endzeiten (oder Sie erstellen eine einzelne Variable mit Endzeiten aus Datums-/Zeitvariablen, wie unten erläutert).
- **Variiert nach Subjekt.** Dies eignet sich für das **Abschneiden auf der linken Seite**, was auch **Verzögerter Eintrag** genannt wird. Wenn Sie beispielsweise Überlebenszeiten für Patienten analysieren, die an einem Rehabilitationsprogramm nach einem Schlaganfall teilnehmen, gehen Sie möglicherweise davon aus, dass das Risiko zum Zeitpunkt des Schlaganfalls beginnt. Wenn Ihre Stichprobe jedoch nur Patienten umfasst, die das Rehabilitationsprogramm überlebt haben, ist die Stichprobe auf der linken Seite in dem Sinne abgeschnitten, dass die beobachteten Überlebenszeiten durch die Rehabilitationslänge "erhöht" sind. Sie können dies berücksichtigen, indem Sie die Zeit angeben, zu der die Patienten die Rehabilitation zu Beginn der Studie beendet haben.

Datums-/Zeitvariablen. Datums-/Zeitvariablen können nicht zum direkten Definieren von Anfang und Ende des Intervalls verwendet werden. Wenn Sie über Datums-/Zeitvariablen verfügen, sollten Sie diese zum Erstellen von Variablen mit Überlebenszeiten verwenden. Wenn kein

Abschneiden auf der linken Seite vorhanden ist, erstellen Sie einfach eine Variable mit Endzeiten auf Grundlage des Unterschieds zwischen dem Eintritt in die Studie und dem Beobachtungsdatum. Wenn ein Abschneiden nach links vorhanden ist, erstellen Sie eine Variable mit Startzeiten auf Grundlage des Unterschieds zwischen dem Datum des Studienbeginns und dem Datum des Eintrags sowie eine Variable mit Endzeiten auf Grundlage des Unterschieds zwischen dem Beginn der Studie und dem Beobachtungsdatum.

Ereignisstatus. Sie benötigen eine Variable, mit der erfasst wird, ob das Subjekt das relevante Ereignis innerhalb des Intervalls erfahren hat. Subjekte, bei denen das Ereignis nicht eingetreten ist, werden rechts zensiert.

Subjekt-ID. Sie können problemlos stückweise konstante, zeitabhängige Einflussvariablen einbeziehen, indem Sie die Beobachtungen für ein einzelnes Subjekt über mehrere Fälle hinweg aufteilen. Wenn Sie beispielsweise die Überlebenszeiten für Patienten nach einem Schlaganfall analysieren, sollten Variablen mit Angaben über ihre Anamnese als Einflussvariablen nützlich sein. Diese Patienten haben im Laufe der Zeit möglicherweise wichtige medizinische Ereignisse erfahren, durch die ihre Anamnese verändert wurde. Die folgende Tabelle zeigt, wie ein solches Daten-Set strukturiert wird: *Patienten-ID* ist die Subjekt-ID, *Endzeit* definiert die beobachteten Intervalle, *Status* erfasst wichtige medizinische Ereignisse und *Vorgeschichte mit Herzinfarkten* und *Vorgeschichte mit Blutungen* sind stückweise konstante, zeitabhängige Einflussvariablen.

Patienten-ID	*Endzeit*	*Status*	*Vorgeschichte mit Herzinfarkten*	*Vorgeschichte mit Blutungen*
1	5	Herzinfarkt	Nein (Generic)	Nein (Generic)
1	7	Blutungen	Ja (Generic)	Nein (Generic)
1	8	Verstorben	Ja (Generic)	Ja (Generic)
2	24	Verstorben	Nein (Generic)	Nein (Generic)
3	8	Herzinfarkt	Nein (Generic)	Nein (Generic)
3	15	Verstorben	Ja (Generic)	Nein (Generic)

Annahmen. Die Fälle in der Datendatei stehen für eine Stichprobe aus einem komplexen Plan, deren Analyse gemäß den Angaben in der Datei erfolgen soll, die im Dialogfeld “Komplexe Stichproben: Plan” ausgewählt wurde.

Üblicherweise werden bei Cox-Regressionsmodellen proportionale Hazards angenommen — d. h., das Verhältnis von Hazards sollte zwischen einzelnen Fällen im Laufe der Zeit nicht variieren. Wenn diese Annahme nicht beibehalten werden kann, müssen Sie dem Modell zeitabhängige Einflussvariablen hinzufügen.

Kaplan-Meier-Analyse. Wenn Sie keine Einflussvariablen auswählen (oder keine ausgewählten Einflussvariablen in das Modell eingeben) und die Produktgrenzenmethode zum Berechnen der Basisüberlebenskurve auf der Registerkarte “Optionen” auswählen, führt die Prozedur eine Kaplan-Meier-Überlebensanalyse durch.

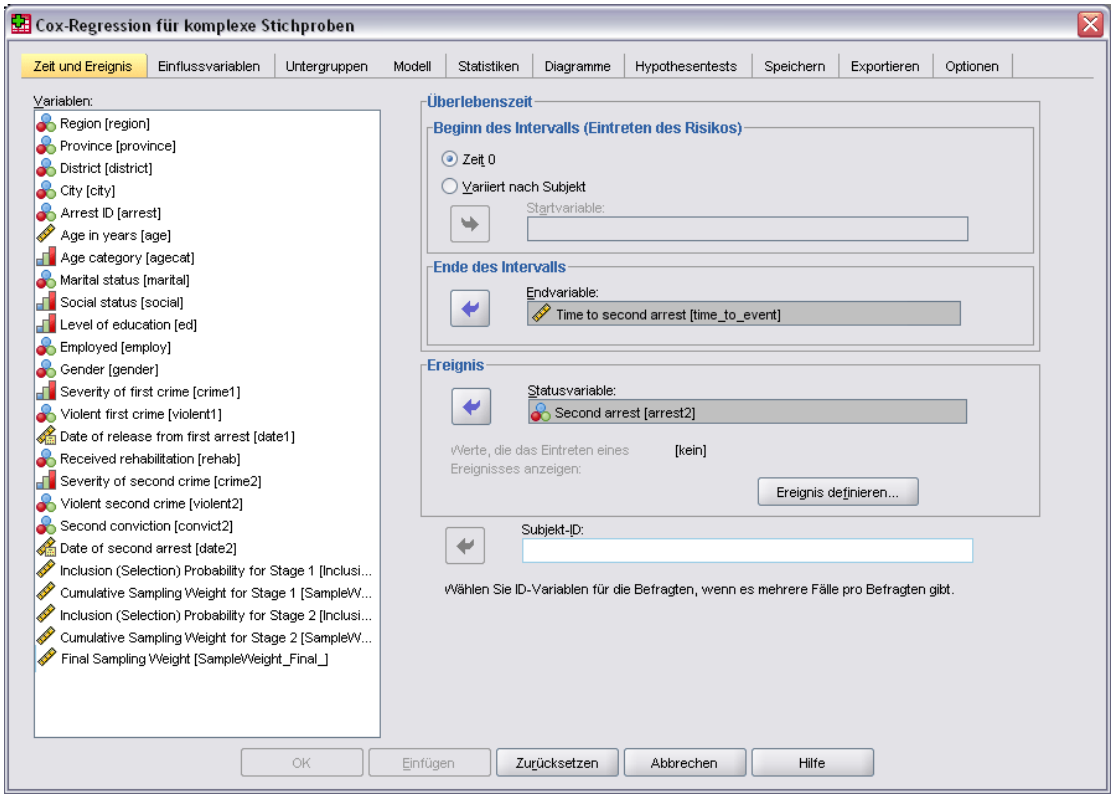
So erstellen Sie die Cox-Regression für komplexe Stichproben

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Komplexe Stichproben
    Cox-Regression…

- Wählen Sie eine Plandatei aus. Optional können Sie eine benutzerdefinierte Datei für die gemeinsamen Wahrscheinlichkeiten auswählen.
- Klicken Sie auf **Weiter**.

Abbildung 12-1
Registerkarte "Zeit und Ereignis", Dialogfeld "Cox-Regression"

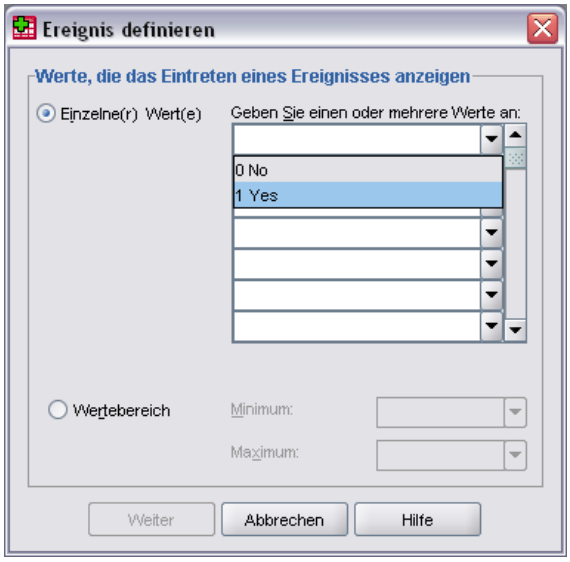


- Geben Sie die Überlebenszeit an, indem Sie die Ein- und Endzeiten aus der Studie auswählen.
- Wählen Sie eine Ereignisstatusvariable aus.
- Klicken Sie auf Ereignis definieren und definieren Sie mindestens einen Ereigniswert.

  Sie können optional auch eine Subjekt-ID auswählen.

Ereignis definieren

Abbildung 12-2
Dialogfeld "Ereignis definieren"

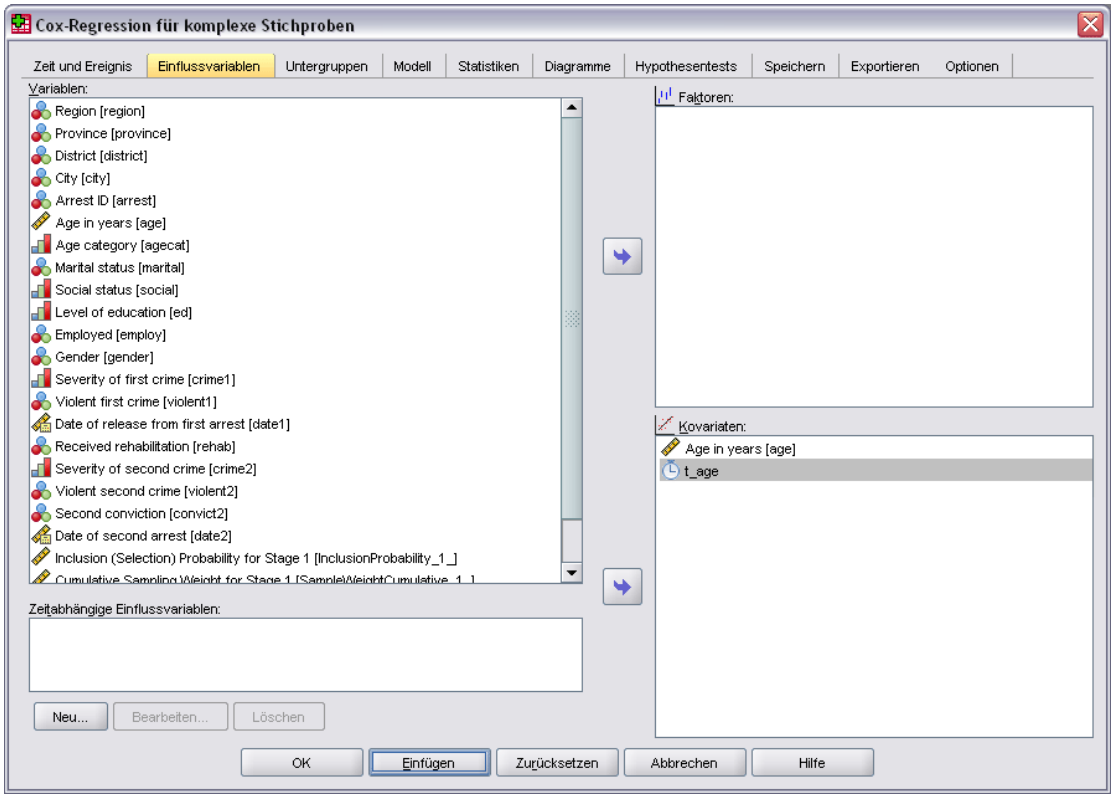


Definieren Sie die Werte, die angeben, dass ein terminales Ereignis eingetreten ist.

- **Einzelne(r) Wert(e).** Geben Sie mindestens einen Wert an, indem Sie ihn in das Gitter eingeben oder aus einer Werteliste mit definierten Wertelabeln auswählen.
- **Wertebereich.** Geben Sie einen Wertebereich an, indem Sie die Mindest- und Maximalwerte eingeben oder aus einer Werteliste mit definierten Wertelabeln auswählen.

Einflussvariablen

Abbildung 12-3
Dialogfeld "Cox-Regression", Registerkarte "Statistiken"



Sie können auf der Registerkarte "Einflussvariablen" die Faktoren und Kovariate zum Erstellen von Modelleffekten angeben.

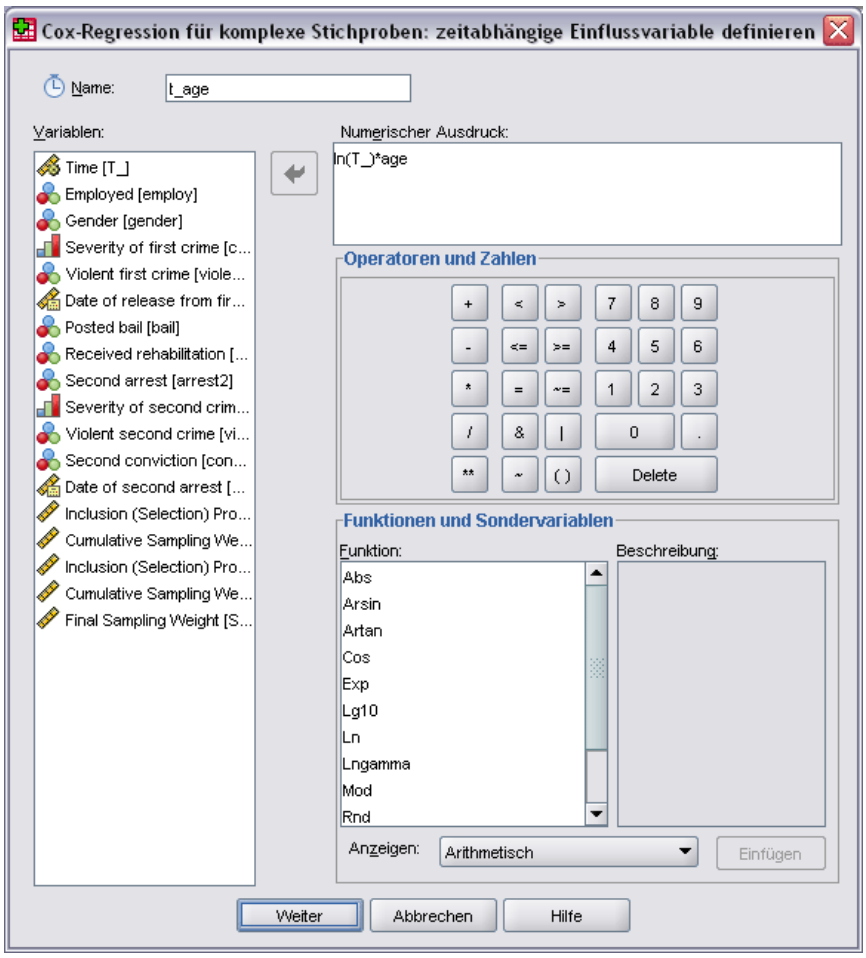
Faktoren. Faktoren sind kategoriale Einflussvariablen; es kann sich dabei um numerische Variablen oder Stringvariablen handeln.

Kovariate. Kovariate sind metrische Einflussvariablen; sie müssen numerisch sein.

Zeitabhängige Einflussvariablen. Es gibt bestimmte Situationen, in denen die Annahme für proportionale Hazards nicht beibehalten werden kann. Dies bedeutet, dass sich Hazard-Verhältnisse über die Zeit ändern können. Die Werte einer (oder mehrerer) Einflussvariablen können zu verschiedenen Zeitpunkten unterschiedlich sein. Sie müssen in solchen Fällen zeitabhängige Einflussvariablen angeben. Für weitere Informationen siehe Definieren einer zeitabhängigen Einflussvariablen auf S. 83. Zeitabhängige Einflussvariablen können als Faktoren oder Kovariate ausgewählt werden.

Definieren einer zeitabhängigen Einflussvariablen

Abbildung 12-4
Cox-Regression: Dialogfeld "Zeitabhängige Einflussvariable definieren"



Sie können im Dialogfeld "Zeitabhängige Einflussvariable definieren" eine Einflussvariable erstellen, die von der integrierten Zeitvariablen *T_* abhängig ist. Sie können diese Variable verwenden, um zeitabhängige Kovariate auf zwei Arten zu definieren.

- Wenn Sie ein erweitertes Cox-Regressionsmodell schätzen möchten, mit dem nicht proportionale Hazards möglich sind, definieren Sie die zeitabhängige Einflussvariable als Funktion der Zeitvariablen *T_* und der fraglichen Kovariate. Ein geläufiges Beispiel wäre das einfache Produkt aus Zeitvariable und Kovariate, aber es können auch komplexere Funktionen festgelegt werden.
- Einige Variablen können zu unterschiedlichen Zeiten unterschiedliche Werte aufweisen, weisen aber kein systematisches Verhältnis zur Zeit auf. Sie müssen in solchen Fällen eine **segmentierte zeitabhängige Einflussvariable** definieren. Dies können Sie mit logischen Ausdrücken erreichen. Logische Ausdrücke nehmen bei wahren Aussagen den Wert 1 und bei falschen Aussagen den Wert 0 an. Mithilfe einer Verkettung von logischen Ausdrücken können Sie die zeitabhängige Einflussvariable aus einem Satz von Messwerten erstellen. Wenn Sie z. B. über vier Wochen einer Studie einmal wöchentlich den Blutdruck gemessen

haben (gekennzeichnet durch *BP1* bis *BP4*), können Sie die zeitabhängige Einflussvariable durch den folgenden Ausdruck definieren: (*T_* < 1) * *BP1* + (*T_* >= 1 & *T_* < 2) * *BP2* + (*T_* >= 2 & *T_* < 3) * *BP3* + (*T_* >= 3 & *T_* < 4) * *BP4*. Beachten Sie, dass bei einem gegebenen Fall genau einer der Ausdrücke in Klammern dem Wert 1 entspricht, alle anderen Ausdrücke in Klammern weisen den Wert 0 auf. Diese Funktion kann folgendermaßen interpretiert werden: Wenn die Zeitspanne kürzer als eine Woche ist, wird *BP1* verwendet, wenn die Zeitspanne länger als eine Woche, aber kürzer als zwei Wochen ist, wird *BP2* verwendet usw.

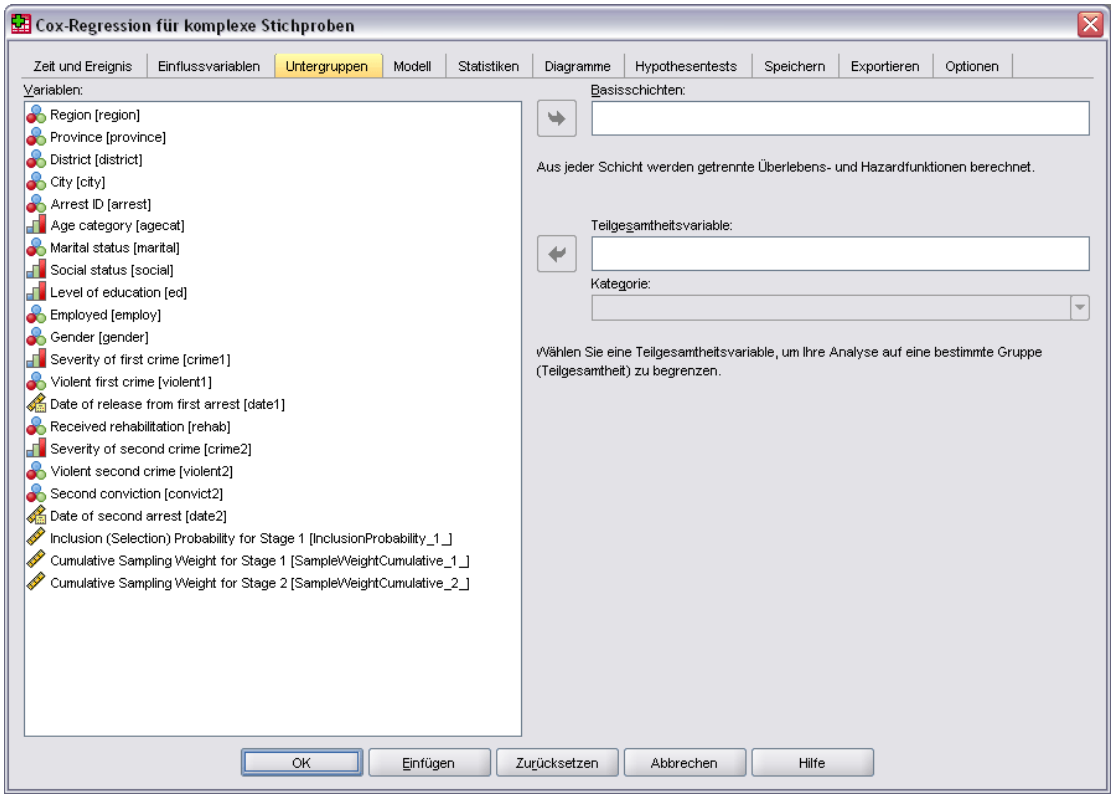
Für segmentierte zeitabhängige Einflussvariablen gilt, dass Fälle mit fehlenden Werten aus der Analyse entfernt werden. Deshalb müssen Sie sicherstellen, dass alle Fälle für jeden gemessenen Zeitpunkt der Einflussvariablen über Werte verfügen. Dies gilt auch für Zeitpunkte, nachdem der Fall aus dem Risikoset entfernt wurde (aufgrund eines Ereignisses oder der Rechtszensur). Diese Werte werden zwar in der Analyse nicht verwendet, müssen aber für SPSS gültige Werte besitzen, damit verhindert wird, dass die Fälle ausgeschlossen werden. Wenn entsprechend der oben angeführten Definition ein Fall beispielsweise in der zweiten Woche zensiert wird, müssen trotzdem Werte für *BP3* und *BP4* vorhanden sein (die Werte können 0 oder eine andere Zahl sein, weil sie in der Analyse nicht verwendet werden).

Anmerkung: Wenn die segmentierte zeitabhängige Einflussvariable innerhalb der Segmente konstant ist, wie in dem Beispiel zum Blutdruck oben, ist es möglicherweise einfacher, die stückweise konstante zeitabhängige Einflussvariable anzugeben, indem Sie die Subjekte über mehrere Fälle hinweg aufteilen. Weitere Informationen finden Sie in der Erörterung zu Subjekt-IDs unter Cox-Regression für komplexe Stichproben auf S. 78.

Sie können die Steuerelemente im Dialogfeld “Zeitabhängige Einflussvariable definieren” verwenden, um den Ausdruck für die zeitabhängige Kovariate zu erstellen, oder sie können diesen direkt in das Textfeld “Numerischer Ausdruck” eingeben. Beachten Sie, dass String-Konstanten in Anführungszeichen oder Apostrophe gesetzt und numerische Konstanten in amerikanischem Format mit einem Punkt als Dezimaltrennzeichen eingegeben werden müssen. Die resultierende Variable erhält den von Ihnen angegebenen Namen und sollte als Faktor oder Kovariate auf der Registerkarte “Einflussvariablen” einbezogen werden.

Untergruppen

Abbildung 12-5
Dialogfeld "Cox-Regression", Registerkarte "Untergruppen"

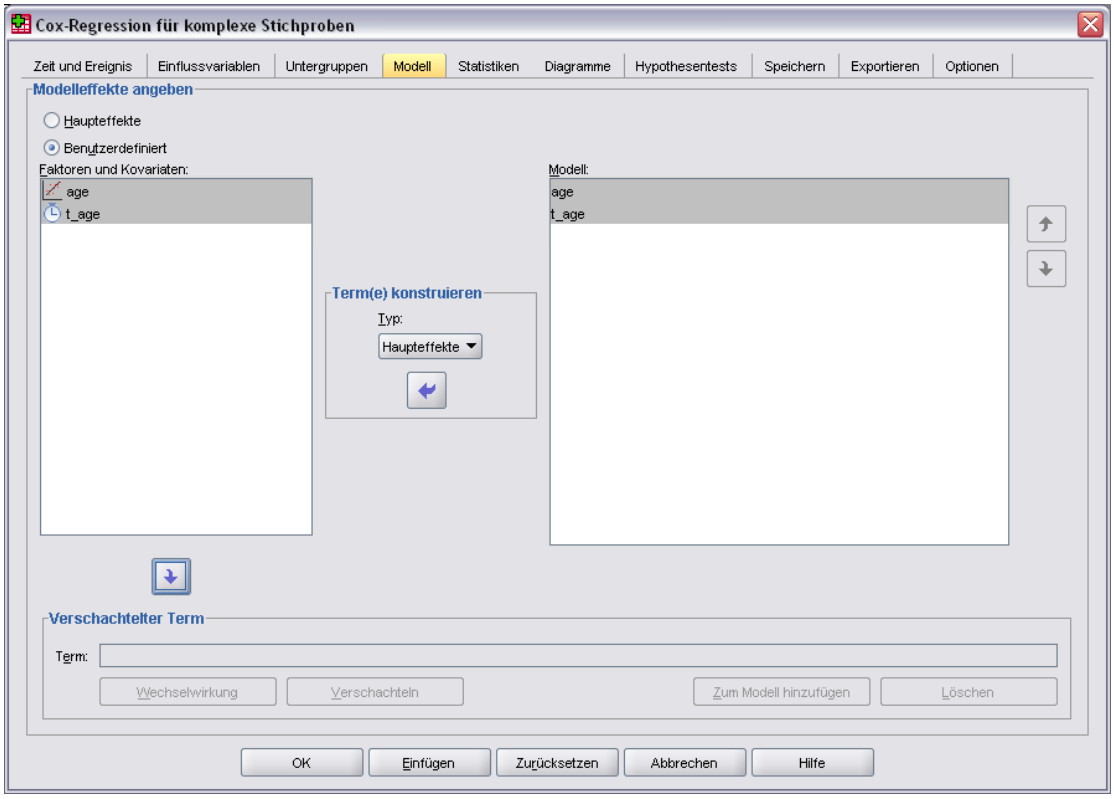


Basisschichten. Für jeden Wert dieser Variablen wird eine getrennte Grundlinien-Hazard- und Überlebensfunktion berechnet, während ein einzelner Satz mit Modellkoeffizienten über Schichten hinweg geschätzt wird.

Teilgesamtheitsvariable. Legen Sie eine Variable fest, um so eine Teilgesamtheit zu definieren. Die Analyse wird ausschließlich für die ausgewählte Kategorie der Teilgesamtheit-Variablen vorgenommen.

Modell

Abbildung 12-6
Dialogfeld "Cox-Regression", Registerkarte "Modell"



Modell-effekte angeben. Standardmäßig erstellt das Verfahren ein Modell mit Haupteffekten unter Verwendung der im Hauptdialogfeld angegebenen Faktoren und Kovariaten. Alternativ können Sie ein benutzerdefiniertes Modell erstellen, das Wechselwirkungseffekte und verschachtelte Terme enthält.

Nicht verschachtelte Terme

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung für alle ausgewählten Variablen erzeugt.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-fach. Hiermit werden alle möglichen 2-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-fach. Hiermit werden alle möglichen 3-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-fach. Hiermit werden alle möglichen 4-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-fach. Hiermit werden alle möglichen 5-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Verschachtelte Terme

In dieser Prozedur können Sie verschachtelte Terme für ein Modell konstruieren. Verschachtelte Terme sind nützlich, um den Effekt von Faktoren oder Kovariaten zu analysieren, deren Werte nicht mit den Stufen eines anderen Faktors interagieren. Eine Lebensmittelkette kann beispielsweise das Kaufverhalten ihrer Kunden in mehreren Filialen untersuchen. Da jeder Kunde nur eine dieser Filialen besucht, kann der Effekt *Kunde* als **verschachtelt innerhalb** des Effekts *Filiale* beschrieben werden.

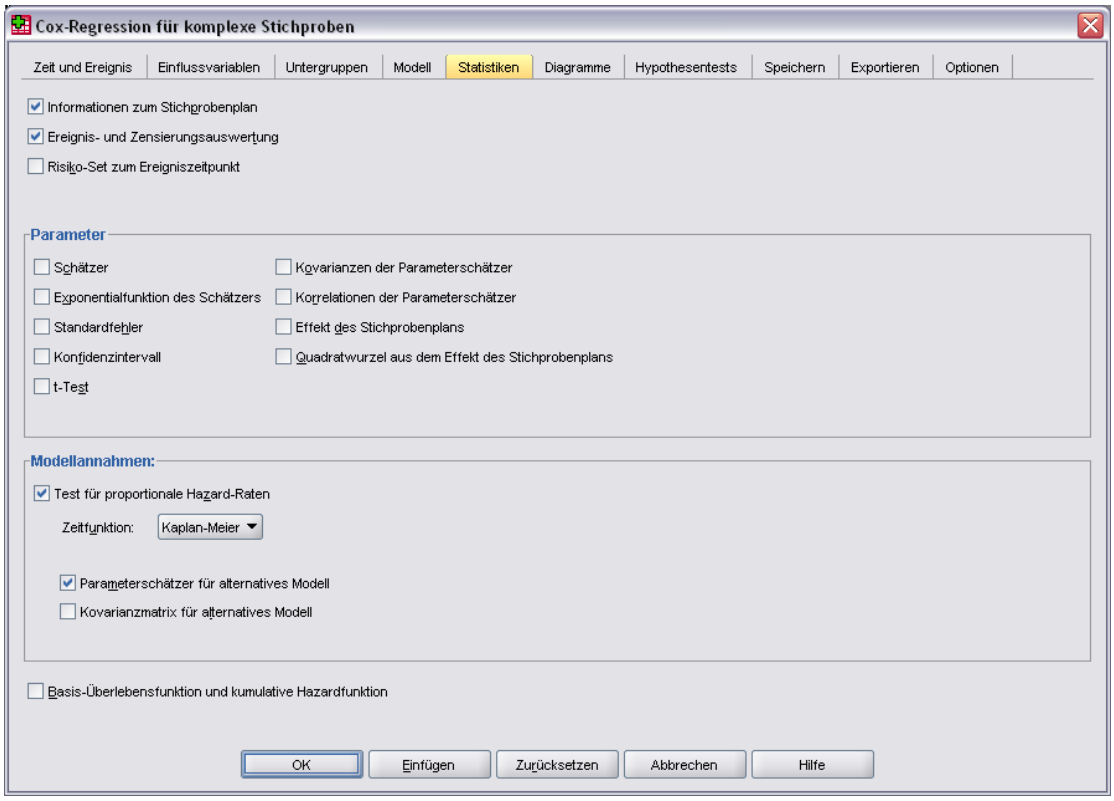
Darüber hinaus können Sie Wechselwirkungseffekte, wie polynomiale Terme mit derselben Kovariaten, einschließen oder dem verschachtelten Term mehrere Verschachtelungsebenen hinzufügen.

Einschränkungen. Für verschachtelte Terme gelten die folgenden Einschränkungen:

- Alle Faktoren innerhalb einer Wechselwirkung müssen eindeutig sein. Dementsprechend ist die Angabe von *A*A* unzulässig, wenn *A* ein Faktor ist.
- Alle Faktoren innerhalb eines verschachtelten Effekts müssen eindeutig sein. Dementsprechend ist die Angabe von *A(A)* unzulässig, wenn *A* ein Faktor ist.
- Effekte dürfen nicht in einer Kovariaten verschachtelt werden. Dementsprechend ist die Angabe von *A(X)* unzulässig, wenn *A* ein Faktor und *X* eine Kovariate ist.

Statistik

Abbildung 12-7
Dialogfeld "Cox-Regression", Registerkarte "Statistiken"



Informationen zum Stichprobenplan. Zeigt eine Zusammenfassung über die Stichprobe (mit ungewichteter Anzahl und Umfang der Grundgesamtheit).

Ereignis- und Zensierungsauswertung. Zeigt zusammenfassende Informationen über die Zahl und den Prozentsatz von zensierten Fällen an.

Risiko-Set zum Ereigniszeitpunkt. Zeigt die Anzahl von Ereignissen und die Anzahl der gefährdeten Fälle für jede Ereigniszeit in jeder Basisschicht an.

Parameter. In dieser Gruppe steuern Sie die Anzeige der Statistiken für die Modellparameter.

- **Schätzer.** Zeigt eine Schätzung der Koeffizienten.
- **Potenzierter Schätzer.** Zeigt die Basis des natürlichen Logarithmus, potenziert mit dem Schätzer der Koeffizienten. Der Schätzer bietet zwar ergiebige Eigenschaften für statische Tests; der potenzierte Schätzer oder exp(B) ist jedoch einfacher zu interpretieren.
- **Standardfehler.** Zeigt den Standardfehler für die einzelnen Koeffizientenschätzer.
- **Konfidenzintervall.** Zeigt ein Konfidenzintervall für die einzelnen Koeffizientenschätzer. Das Konfidenzniveau für das Intervall wird im Dialogfeld "Optionen" festgelegt.
- **T-Test.** Zeigt je einen *t*-Test für die einzelnen Koeffizientenschätzer. Die Nullhypothese der Tests liegt vor, wenn der Koeffizient den Wert 0 aufweist.

- **Kovarianzen der Parameterschätzer.** Zeigt eine Schätzung der Kovarianzmatrix für die Modellkoeffizienten.
- **Korrelationen der Parameterschätzer.** Zeigt eine Schätzung der Korrelationsmatrix für die Modellkoeffizienten.
- **Effekt des Stichprobenplans.** Das Verhältnis der Varianz des Schätzers zur Varianz unter der Annahme, dass es sich bei der Stichprobe um eine einfache Zufallsstichprobe handelt. Ein Maß für den Effekt eines komplexen Stichprobenplans; kleinere Werte weisen auf größere Effekte hin.
- **Quadratwurzel aus dem Effekt des Stichprobenplans.** Dies ist ein Maß für den Effekt der Angabe eines komplexen Plans. Je stärker der Wert von 1 abweicht, desto größer ist der Effekt.

Modellannahmen. Mit dieser Gruppe können Sie einen Test der proportionalen Hazard-Annahme durchführen. Der Test vergleicht das angepasste Modell mit einem alternativen Modell mit zeitabhängigen Einflussvariablen: *x**_TF* für jede Einflussvariable *x*, wobei *_TF* die angegebene Zeitfunktion ist.

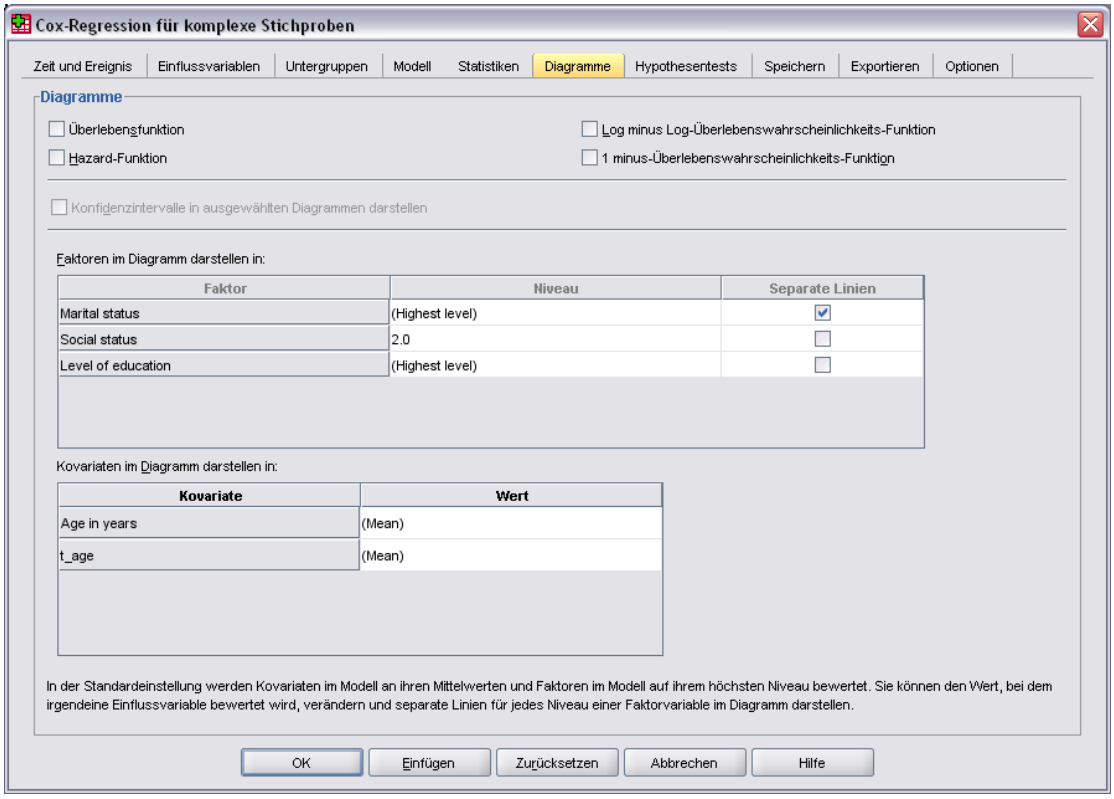
- **Zeitfunktion.** Gibt *_TF* für das alternative Modell an. Für die Funktion **Identität**: *_TF=T_*. Für die Funktion **Log**: *_TF*=log(*T_*). Für **Kaplan-Meier**: *_TF*=1−S_{KM}(*T_*), wobei $S_{KM}(.)$ die Kaplan-Meier-Schätzung der Überlebensfunktion ist. Für **Rang**: *_TF* ist die Rangreihenfolge von *T_* zwischen den beobachteten Endzeiten.
- **Parameterschätzer für alternatives Modell.** Zeigt die Schätzung, den Standardfehler und das Konfidenzintervall für jeden Parameter im alternativen Modell an.
- **Kovarianzmatrix für alternatives Modell.** Zeigt die Matrix von geschätzten Kovarianzen zwischen Parametern im alternativen Modell an.

Basis-Überlebensfunktion und kumulative Hazard-Funktion. Zeigt die Basis-Überlebens- sowie die kumulative Basis-Hazard-Funktion zusammen mit den zugehörigen Standardfehlern an.

Anmerkung: Wenn auf der Registerkarte “Einflussvariablen” definierte zeitabhängige Einflussvariablen Bestandteil des Modells sind, ist diese Option nicht verfügbar.

Diagramme

Abbildung 12-8
Dialogfeld "Cox-Regression", Registerkarte "Diagramme"



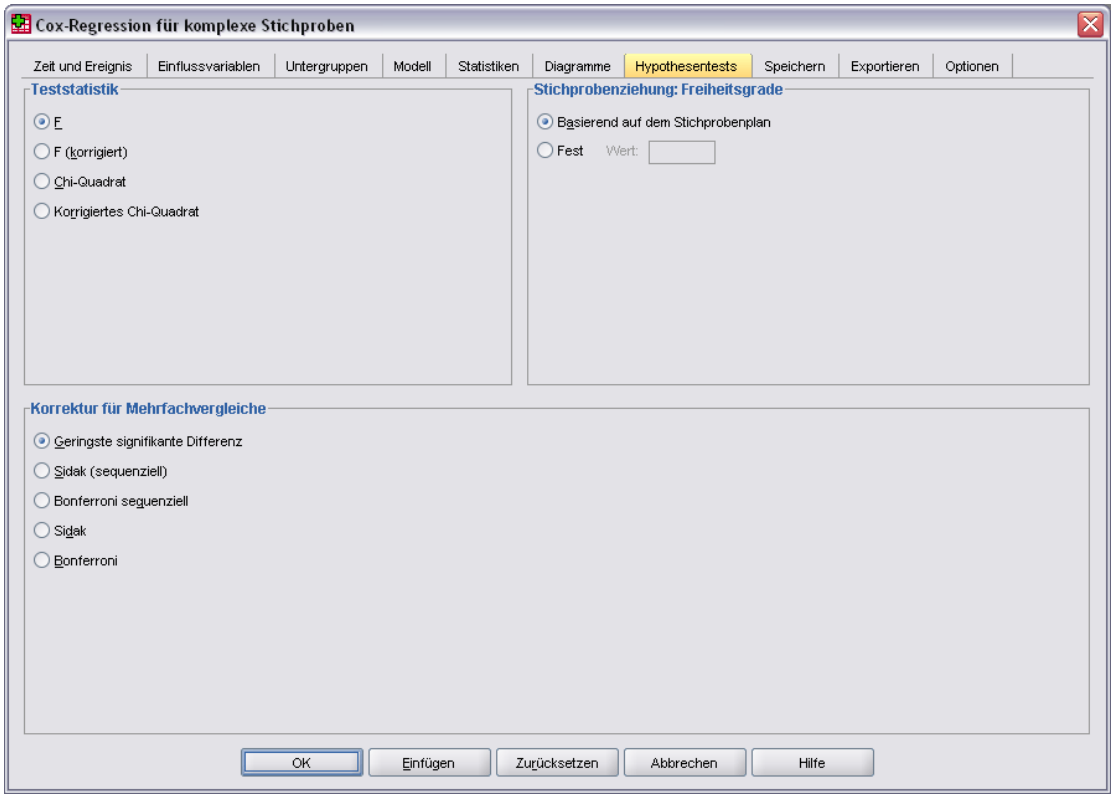
Sie können auf der Registerkarte "Diagramme" Diagramme der Hazard-Funktion, der Überlebensfunktion, von Log-minus-Log der Überlebensfunktion und von 1- der Überlebensfunktion abfragen. Sie können auch Diagramme der Konfidenzintervalle für die angegebenen Funktionen erstellen. Das Konfidenzniveau wird auf der Registerkarte "Optionen" festgelegt.

Einflussvariablenmuster. Sie können ein Muster mit Einflussvariablen angeben, die für die angefragten Diagramme und die exportierte Überlebensdatei auf der Registerkarte "Exportieren" verwendet werden soll. Beachten Sie, dass diese Optionen nicht verfügbar sind, wenn auf der Registerkarte "Einflussvariablen" definierte zeitabhängige Einflussvariablen Bestandteil des Modells sind.

- **Faktoren im Diagramm darstellen in.** Standardmäßig wird jeder Faktor auf seiner jeweiligen höchsten Stufe ausgewertet. Geben Sie bei Bedarf eine andere Stufe ein oder wählen Sie sie aus. Sie können alternativ auch ein Diagramm mit separaten Linien für jede Stufe eines einzelnen Faktors erstellen, indem Sie das Kontrollkästchen für den jeweiligen Faktor aktivieren.
- **Kovariaten im Diagramm darstellen in.** Jede Kovariate wird bei ihrem Mittelwert ausgewertet. Geben Sie bei Bedarf eine andere Stufe ein oder wählen Sie sie aus.

Hypothesentests

Abbildung 12-9
Dialogfeld "Cox-Regression", Registerkarte "Hypothesentests"



Teststatistik. In dieser Gruppe können Sie den Typ der Statistik zum Testen der Hypothesen festlegen. Die folgenden Optionen stehen zur Auswahl: *F*, *F* (korrigiert), "Chi-Quadrat" und "Korrigiertes Chi-Quadrat".

Stichprobenziehung: Freiheitsgrade. In dieser Gruppe steuern Sie die Freiheitsgrade im Stichprobenplan, mit denen die *p*-Werte für alle Teststatistiken berechnet werden. Dient der Stichprobenplan als Grundlage, ist dieser Wert die Differenz zwischen der Anzahl der primären Stichprobeneinheiten und der Anzahl der Schichten in der ersten Stufe der Stichproben. Alternativ können Sie benutzerdefinierte Freiheitsgrade festlegen; geben Sie hierzu eine positive Ganzzahl ein.

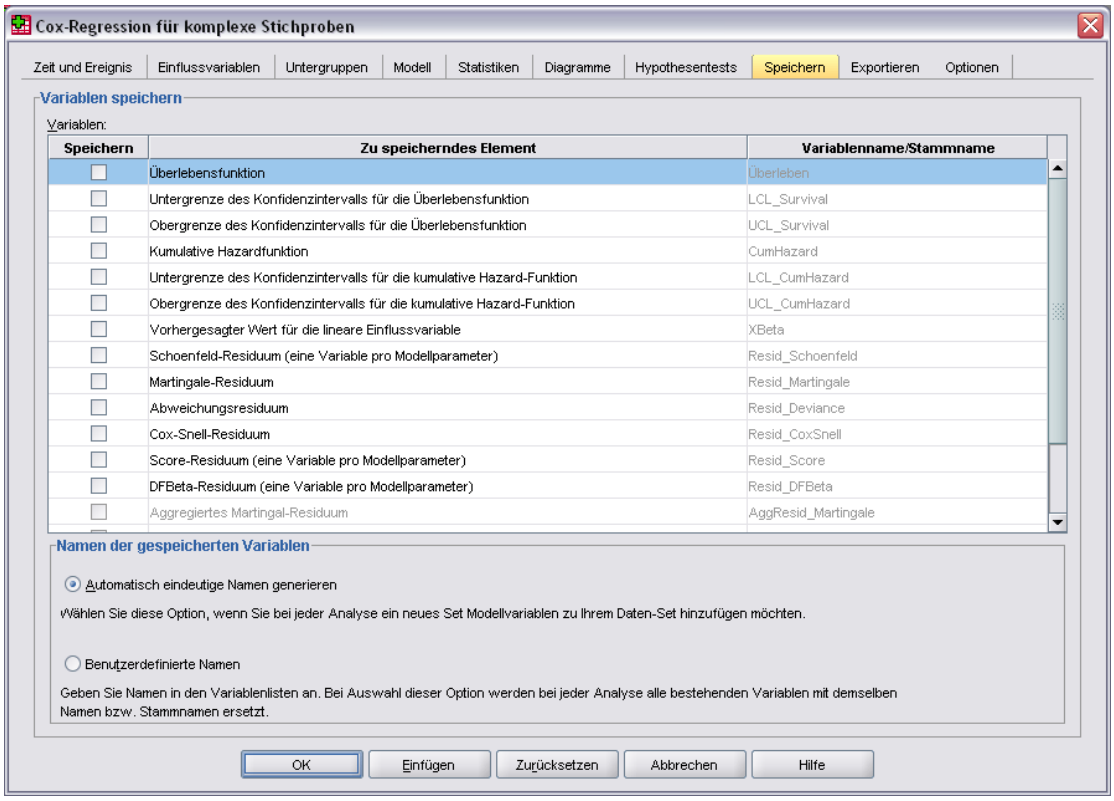
Korrektur für Mehrfachvergleiche. Bei der Durchführung von Hypothesentests mit mehreren Kontrasten kann das Gesamtsignifikanzniveau mithilfe der Signifikanzniveaus der eingeschlossenen Kontraste angepasst werden.. In dieser Gruppe können Sie die Anpassungs-/Korrekturmethode auswählen.

- **Geringste signifikante Differenz.** Diese Methode steuert nicht die Gesamtwahrscheinlichkeit, dass Hypothesen abgelehnt werden, bei denen einige lineare Kontraste von den Werten einer Nullhypothese abweichen.

- **Sidak (sequenziell).** Hierbei handelt es sich um ein sequenzielles schrittweises Sidak-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen anbelangt, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.
- **Bonferroni sequenziell.** Hierbei handelt es sich um ein sequenzielles schrittweises Bonferroni-Verfahren, das deutlich weniger konservativ ist, was die Ablehnung einzelner Hypothesen anbelangt, aber dennoch dasselbe allgemeine Signifikanzniveau beibehält.
- **Sidak.** Dieses Verfahren liefert engere Grenzen als der Bonferroni-Ansatz.
- **Bonferroni.** Dieses Verfahren passt das empirische Signifikanzniveau der Tatsache an, dass mehrere Kontraste getestet werden.

Speichern

Abbildung 12-10
Dialogfeld "Cox-Regression", Registerkarte "Speichern"



Variablen speichern. Mit dieser Gruppe können Sie modellbezogene Variablen im aktiven Daten-Set speichern, um diese später für die Diagnose und Berichterstellung von Ergebnissen zu verwenden. Beachten Sie, dass keine dieser Variablen verfügbar ist, wenn zeitabhängige Einflussvariablen Bestandteil des Modells sind.

- **Überlebensfunktion.** Speichert die Überlebenswahrscheinlichkeit (den Wert der Überlebensfunktion) zur beobachteten Zeit sowie Einflussvariablenwerte für jeden Fall.

- **Untergrenze des Konfidenzintervalls für die Überlebensfunktion.** Speichert die Untergrenze des Konfidenzintervalls für die Überlebensfunktion zur beobachteten Zeit sowie Einflussvariablenwerte für jeden Fall.
- **Obergrenze des Konfidenzintervalls für die Überlebensfunktion.** Speichert die Obergrenze des Konfidenzintervalls für die Überlebensfunktion zur beobachteten Zeit sowie Einflussvariablenwerte für jeden Fall.
- **Kumulative Hazard-Funktion.** Speichert den kumulativen Hazard oder −ln(survival) zur beobachteten Zeit sowie Einflussvariablenwerte für jeden Fall.
- **Untergrenze des Konfidenzintervalls für die kumulative Hazard-Funktion.** Speichert die Untergrenze des Konfidenzintervalls für die kumulative Hazard-Funktion zur beobachteten Zeit sowie Einflussvariablenwerte für jeden Fall.
- **Obergrenze des Konfidenzintervalls für die kumulative Hazard-Funktion.** Speichert die Obergrenze des Konfidenzintervalls für die kumulative Hazard-Funktion zur beobachteten Zeit sowie Einflussvariablenwerte für jeden Fall.
- **Vorhergesagter Wert für die lineare Einflussvariable.** Speichert die lineare Kombination von im Referenzwert berichtigten Regressionskoeffizienten der Einflussvariablenzeiten. Die lineare Einflussvariable ist das Verhältnis von Hazard-Funktion zu Basis-Hazard. Bei dem proportionalen Hazard-Modell ist dieser Wert im Laufe der Zeit konstant.
- **Schoenfeld-Residuum.** Für jeden nicht zensierten Fall und jeden nicht redundanten Parameter im Modell bedeutet das Schoenfeld-Residuum den Unterschied zwischen der beobachteten Einflussvariablen, die mit dem Modellparameter verknüpft ist, und dem erwarteten Wert der Einflussvariablen für Fälle im Risikoset zur beobachteten Ereigniszeit. Schoenfeld-Residuen können verwendet werden, um die Beurteilung der proportionalen Hazard-Annahme zu unterstützen, beispielsweise sollten bei einer Einflussvariable x Diagramme der Schoenfeld-Residuen für die zeitabhängige Einflussvariable x*ln($T\_$) im Vergleich zur Zeit eine horizontale Linie bei 0 zeigen, wenn proportionale Hazards sich als richtig erweisen. Für jeden nicht redundanten Parameter im Modell wird eine separate Variable gespeichert. Schoenfeld-Residuen werden nur für nicht zensierte Fälle berechnet.
- **Martingale-Residuum.** Das Martingale-Residuum bedeutet für jeden Fall den Unterschied zwischen der beobachteten Zensierung (0 bei Zensierung, 1 ohne Zensierung) und der Erwartung eines Ereignisses während der Beobachtungszeit.
- **Abweichungsresiduum.** Abweichungsresiduen sind Martingale-Residuen, die "angepasst" wurden, um symmetrischer bei 0 zu erscheinen. Diagramme von Abweichungsresiduen gegen Einflussvariablen sollten keine Muster aufweisen.
- **Cox-Snell-Residuum.** Das Cox-Snell-Residuum ist für jeden Fall die Ausnahme eines Ereignisses während der Beobachtungszeit oder die beobachtete Zensierung minus dem Martingale-Residuum.
- **Score-Residuum.** Das Score-Residuum ist für jeden Fall und jeden nicht redundanten Parameter im Modell der Anteil der ersten Ableitung des Falls an die Pseudo-Likelihood. Für jeden nicht redundanten Parameter im Modell wird eine separate Variable gespeichert.
- **DFBeta-Residuum.** Das DFBeta-Residuum schätzt für jeden Fall und jeden nicht redundanten Parameter im Modell die Änderung des Werts der Parameterschätzung, wenn der Fall aus dem Modell entfernt wird. Fälle mit relativ großen DFBeta-Residuen üben möglicherweise

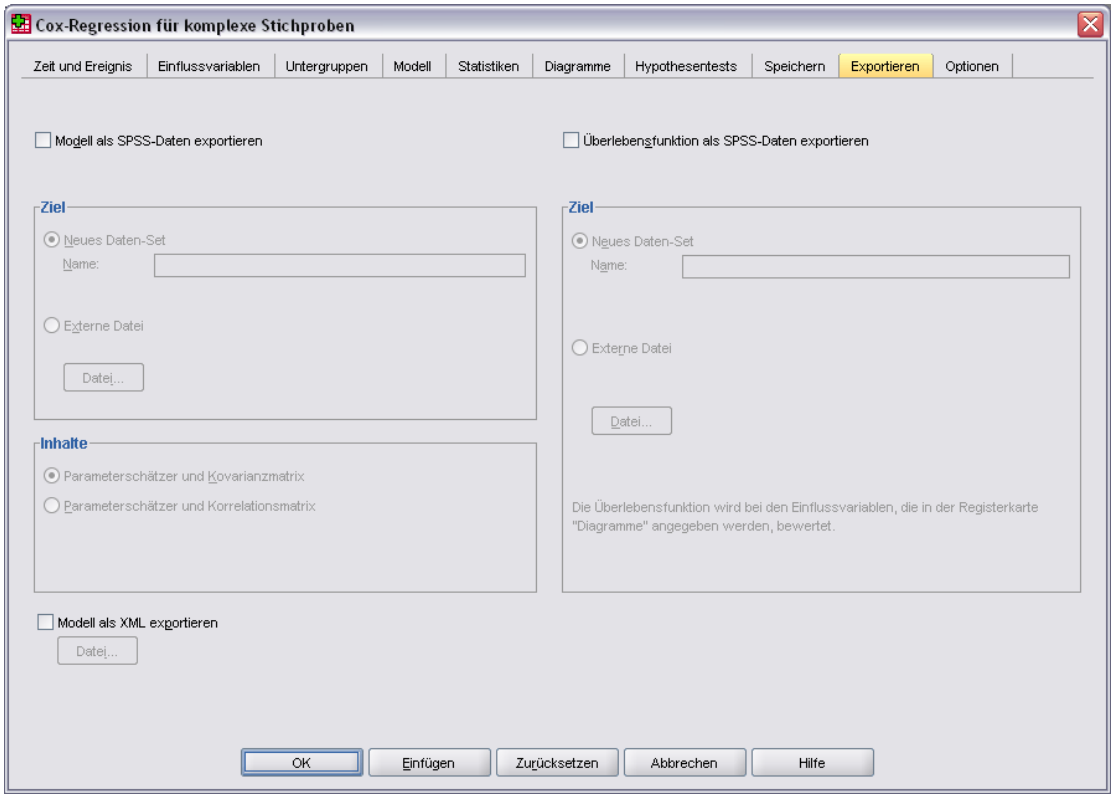
einen übermäßigen Einfluss auf die Analyse aus. Für jeden nicht redundanten Parameter im Modell wird eine separate Variable gespeichert.

- **Aggregierte Residuen.** Wenn mehrere Fälle ein einzelnes Subjekt darstellen, ist das aggregierte Residuum für ein Subjekt einfach die Summe der entsprechenden Fall-Residuen für alle Fälle, die zu demselben Subjekt gehören. Bei dem Schoenfeld-Residuum ist die aggregierte Version identisch mit der nicht aggregierten Version, da das Schoenfeld-Residuum nur für nicht zensierte Fälle definiert wird. Diese Residuen sind nur dann verfügbar, wenn eine Subjekt-ID auf der Registerkarte "Zeit und Ereignis" angegeben wird.

Namen der gespeicherten Variablen. Durch eine automatische Generierung von Namen wird sichergestellt, dass Ihre Arbeit nicht verloren geht. Mit benutzerdefinierten Namen können Sie Ergebnisse aus früheren Durchgängen verwerfen/ersetzen, ohne zuerst die gespeicherten Variablen im Daten-Editor löschen zu müssen.

Exportieren

Abbildung 12-11
Dialogfeld "Cox-Regression", Registerkarte "Exportieren"



Modell als SPSS-Daten exportieren. Schreibt ein SPSS-Daten-Set, das die Parameter-Korrelations- oder -Kovarianzmatrix enthält (mit Parameterschätzern, Standardfehlern, Signifikanzwerten und Freiheitsgraden). Die Reihenfolge der Variablen in der Matrixdatei lautet wie folgt:

- **rowtype_.** Nimmt folgende Werte (und Wertelabel) an: COV (Kovarianzen), CORR (Korrelationen), EST (Parameterschätzer), SE (Standardfehler), SIG (Signifikanzniveaus) und DF (Freiheitsgrade des Stichprobenplans). Es gibt einen separaten Fall mit dem Zeilentyp COV (bzw. CORR) für jeden Modellparameter sowie einen separaten Fall für jeden der anderen Zeilentypen.
- **varname_.** Nimmt für die Zeilentypen COV bzw. CORR die Werte P1, P2, ... an, was einer geordneten Liste aller geschätzten Modellparameter entspricht, mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Für alle anderen Zeilentypen sind die Zellen leer.
- **P1, P2, ...** Diese Variablen entsprechen einer geordneten Liste aller Modellparameter mit Wertelabels, die den in der Tabelle der Parameterschätzer angezeigten Parameter-Strings entsprechen. Die Werte hängen jeweils vom Zeilentyp ab. Bei redundanten Parametern sind alle Kovarianzen auf 0 gesetzt, die Korrelationen sind auf den systemdefiniert fehlenden Wert gesetzt, alle Parameterschätzer sind auf 0 gesetzt und alle Standardfehler, Signifikanzniveaus und die Freiheitsgrade der Residuen sind auf den systemdefiniert fehlenden Wert gesetzt.

Anmerkung: Diese Datei ist nicht unmittelbar für weitere Analysen in anderen Prozeduren verwendbar, bei denen eine Matrixdatei eingelesen wird, es sei denn, diese Prozeduren akzeptieren alle hier exportierten Zeilentypen.

Überlebensfunktion als SPSS-Daten exportieren. Schreibt ein SPSS-Daten-Set mit der Überlebensfunktion, dem Standardfehler der Überlebensfunktion, den Ober- und Untergrenzen des Konfidenzintervalls der Überlebensfunktion sowie der kumulativen Hazard-Funktion für jeden Fehler oder jede Ereigniszeit, die auf der Registerkarte “Diagramm” an der Grundlinie und den Einflussvariablenwerten ausgewertet wird. Die Reihenfolge der Variablen in der Matrixdatei lautet wie folgt:

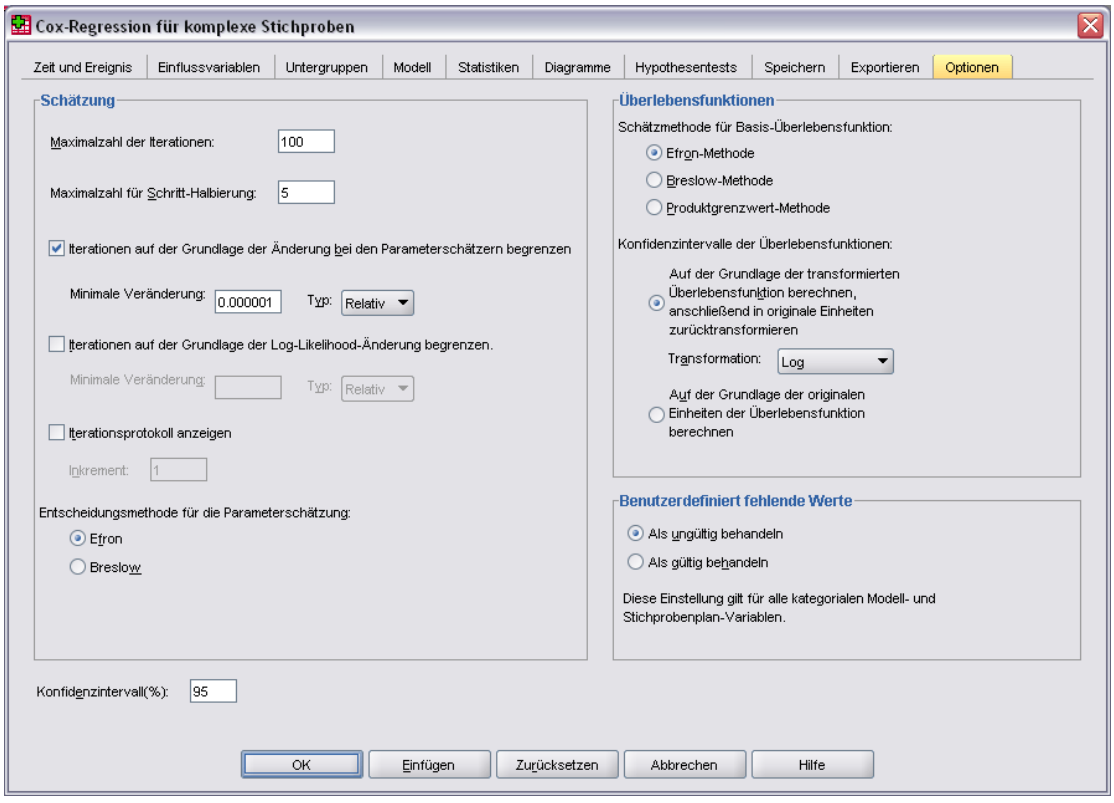
- **Basisschichtvariable.** Für jeden Wert der Schichtvariablen werden separate Überlebenstabellen erstellt.
- **Überlebenszeitvariable.** Die Ereigniszeit. Für jede eindeutige Ereigniszeit wird ein separater Fall erstellt.
- **Sur_0, LCL_Sur_0, UCL_Sur_0.** Basisüberlebensfunktion und die Ober- und Untergrenzen des zugehörigen Konfidenzintervalls.
- **Sur_R, LCL_Sur_R, UCL_Sur_R.** Überlebensfunktion, die beim Muster “Referenz” ausgewertet wird (siehe die Tabelle mit Musterwerten in der Ausgabe) sowie die Ober- und Untergrenzen des zugehörigen Konfidenzintervalls.
- **Sur_#.#, LCL_Sur_#.#, UCL_Sur_#.#, …** Überlebensfunktion, die bei jedem auf der Registerkarte “Diagramme” angegebenen Einflussvariablenmuster ausgewertet wird, sowie die Ober- und Untergrenze der zugehörigen Konfidenzintervalle. Siehe die Tabelle mit Musterwerten in der Ausgabe, um eine Übereinstimmung mit den Mustern mit der Zahl #.# zu erzielen.
- **Haz_0, LCL_Haz_0, UCL_Haz_0.** Kumulative Basis-Hazard-Funktion sowie die Ober- und Untergrenzen des zugehörigen Konfidenzintervalls.

- **Haz_R, LCL_Haz_R, UCL_Haz_R.** Kumulative Hazard-Funktion, die beim Muster "Referenz" ausgewertet wird (siehe die Tabelle mit Musterwerten in der Ausgabe), sowie die Ober- und Untergrenzen des zugehörigen Konfidenzintervalls.
- **Haz_#.#, LCL_Haz_#.#, UCL_Haz_#.#, …** Kumulative Hazard-Funktion, die bei jedem auf der Registerkarte "Diagramme" angegebenen Einflussvariablenmuster ausgewertet wird, sowie die Ober- und Untergrenze der zugehörigen Konfidenzintervalle. Siehe die Tabelle mit Musterwerten in der Ausgabe, um eine Übereinstimmung mit den Mustern mit der Zahl #.# zu erzielen.

Modell als XML exportieren. Speichert alle zum Vorhersagen der Überlebensfunktion erforderlichen Informationen, einschließlich von Parameterschätzern und der Basis-Überlebensfunktion im XML-(PMML-)Format. SmartScore und SPSS Server (gesondertes Produkt) können anhand dieser Modelldatei die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.

Optionen

Abbildung 12-12
Dialogfeld "Cox-Regression," Registerkarte "Optionen"



Schätzung. Diese Steuerelemente geben Kriterien für die Schätzung von Regressionskoeffizienten an.

- **Maximalzahl der Iterationen.** Dies ist die maximale Anzahl der Iterationen, die im Algorithmus vorgenommen werden. Geben Sie eine nichtnegative Ganzzahl an.

- **Maximalzahl für Schritt-Halbierung.** Bei jeder Iteration wird die Schrittgröße um den Faktor 0,5 reduziert, bis die Log-Likelihood ansteigt oder die Maximalzahl für die Schritt-Halbierung erreicht ist. Geben Sie eine positive Ganzzahl ein.
- **Iterationen auf der Grundlage der Änderung bei den Parameterschätzern begrenzen.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei den Parameterschätzern unter dem angegebenen (positiven) Wert liegt.
- **Iterationen auf der Grundlage der Log-Likelihood-Änderung begrenzen.** Mit dieser Option wird der Algorithmus nach einer Iteration angehalten, bei der die absolute oder relative Änderung bei der Log-Likelihood-Funktion unter dem angegebenen (positiven) Wert liegt.
- **Iterationsprotokoll anzeigen.** Zeigt das Iterationsprotokoll für Parameterschätzer und Log-Likelihood an und druckt die letzte Auswertung der Änderung von Parameterschätzern und der Pseudo-Log-Likelihood an. Die Tabelle mit dem Iterationsprotokoll druckt alle *n* Iterationen ausgehend von der *0*. Iteration (die Anfangsschätzungen), wobei *n* der Wert des Inkrements ist. Wenn das Iterationsprotokoll angefordert wird, wird die letzte Iteration stets angezeigt, unabhängig von *n*.
- **Entscheidungsmethode für Parameterschätzer.** Wenn gebundene beobachtete Fehlerzeiten vorhanden sind, wird eine dieser Methoden zum Auflösen der Bindungen verwendet. Die Efron-Methode ist rechenaufwendiger.

Überlebensfunktionen. Diese Steuerelemente geben Kriterien für Berechnungen mit der Überlebensfunktion an.

- **Schätzmethode für Basis-Überlebensfunktion.** Mit der **Breslow-** (oder Nelson-Aalan- bzw. empirischen) Methode wird der kumulative Basis-Hazard durch eine nicht absteigende Schrittfunktion mit Schritten zu den beobachteten Fehlerzeiten geschätzt und anschließend die Basis-Überlebensfunktion nach dem Bezugsüberleben=exp(−cumulative hazard) berechnet. Die **Efron**-Methode ist rechenintensiver und wird auf die Breslow-Methode reduziert, wenn keine Bindungen vorhanden sind. Mit der **Produktgrenzen**-Methode wird das Basis-Überleben durch eine nicht aufsteigende rechte stetige Funktion geschätzt. Wenn im Modell keine Einflussvariablen vorhanden sind, wird diese Methode auf die Kaplan-Meier-Schätzung reduziert.
- **Konfidenzintervalle der Überlebensfunktionen.** Das Konfidenzintervall kann auf drei Arten berechnet werden: In ursprünglichen Einheiten, über eine Log-Transformation oder über eine Minus-Log-Transformation. Es wird nur durch die Log-Minus-Log-Transformation gewährleistet, dass die Grenzen des Konfidenzintervalls zwischen 0 und 1 liegen, im Allgemeinen scheint die Log-Transformation am besten zu funktionieren.

Benutzerdefiniert fehlende Werte. Alle Variablen müssen über gültige Werte für einen Fall verfügen, damit dieser in die Analyse einbezogen werden kann. Mit diesen Steuerelementen können Sie entscheiden, ob benutzerdefiniert fehlende Werte bei kategorialen Modellen (einschließlich Faktoren-, Ereignis-, Einheiten- und Teilgesamtheitsvariablen) und bei Stichprobenplan-Variablen als gültig behandelt werden.

Konfidenzintervall (%). Dies ist die Konfidenzintervallstufe, die für Koeffizientenschätzungen, potenzierte Koeffizientenschätzungen, Überlebensfunktionsschätzungen und Schätzungen für kumulative Hazard-Funktionen verwendet wird. Geben Sie einen Wert größer oder gleich 0 und kleiner als 100 an.

Zusätzliche Funktionen des CSCOXREG-Befehls

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Führen Sie benutzerdefinierte Hypothesentests durch (mithilfe des Unterbefehls `CUSTOM` und `/PRINT LMATRIX`).
- Toleranzspezifikation (mithilfe von `/CRITERIA SINGULAR`).
- Allgemein schätzbare Funktionen (mithilfe von `/PRINT GEF`).
- Mehrere Einflussvariablenmuster (mithilfe von mehreren Unterbefehlen `PATTERN`).
- Maximale Anzahl von gespeicherten Variablen, wenn ein Stammname angegeben wird (mithilfe des Unterbefehls `SAVE`). Das Dialogfeld berücksichtigt den `CSCOXREG`-Standard von 25 Variablen.

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Teil II: Beispiele

Kapitel
13

Stichprobenassistent für komplexe Stichproben

Der Stichprobenassistent führt Sie durch die Schritte zum Erstellen, Bearbeiten bzw. Ausführen einer Stichprobenplan-Datei. Vor der Verwendung des Assistenten sollten Sie über eine klar umrissene Ziel-Grundgesamtheit und eine Liste der Stichprobeneinheiten verfügen und einen geeigneten Stichprobenplan im Kopf haben.

Ziehen einer Stichprobe aus einem vollständigen Stichprobenrahmen

Eine bundesstaatliche Behörde ist damit beauftragt, gerechte Vermögenssteuern in den verschiedenen Counties zu gewährleisten. Die Steuern beruhen auf der Schätzung des Immobilienwerts. Daher möchte die Behörde eine Stichprobe der Immobilien in den einzelnen Counties untersuchen, um sicherzugehen, dass die Akten jedes County gleichermaßen auf dem neuesten Stand sind. Die Ressourcen für die Gewinnung aktueller Schätzungen sind jedoch begrenzt, daher ist ein sinnvoller Einsatz der vorhandenen Ressourcen besonders wichtig. Die Behörde entscheidet sich für die Anwendung eines Verfahrens mit komplexen Stichproben zur Auswahl einer Stichprobe der Immobilien.

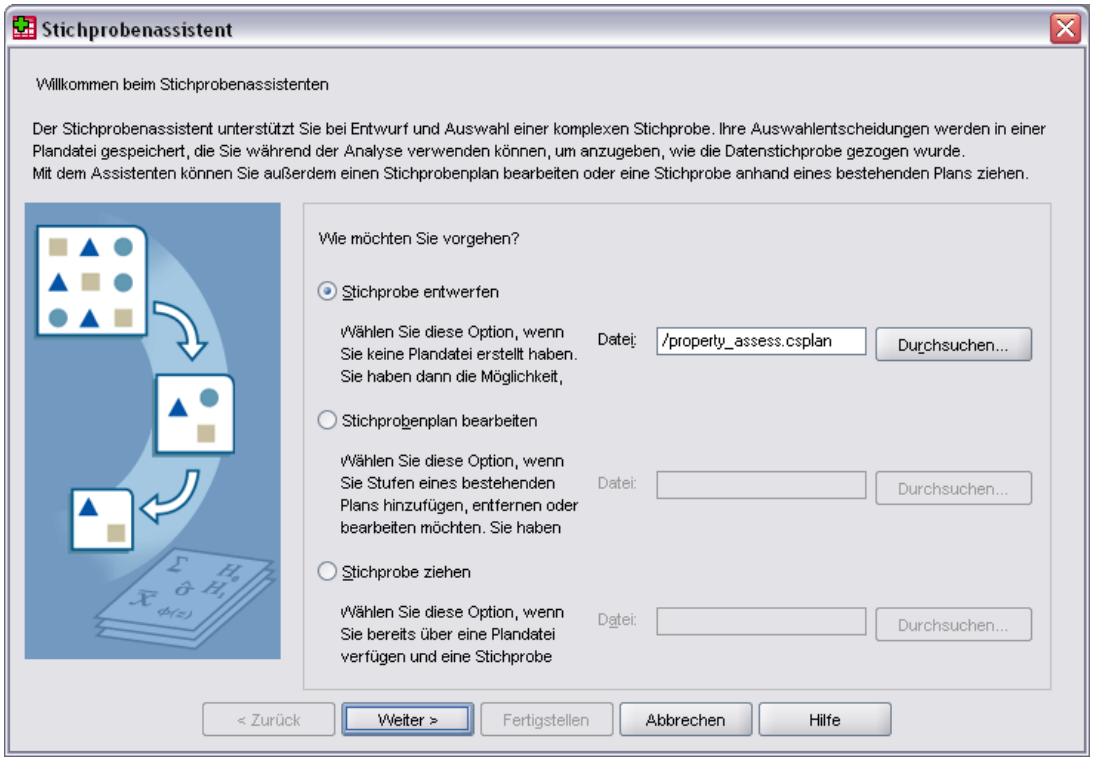
Eine Liste der Immobilien finden Sie in *property_assess_cs.sav*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Verwenden Sie den Stichprobenassistenten für komplexe Stichproben, um eine Stichprobe zu ziehen.

Verwendung des Assistenten

- Um den Stichprobenassistenten für komplexe Stichproben durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

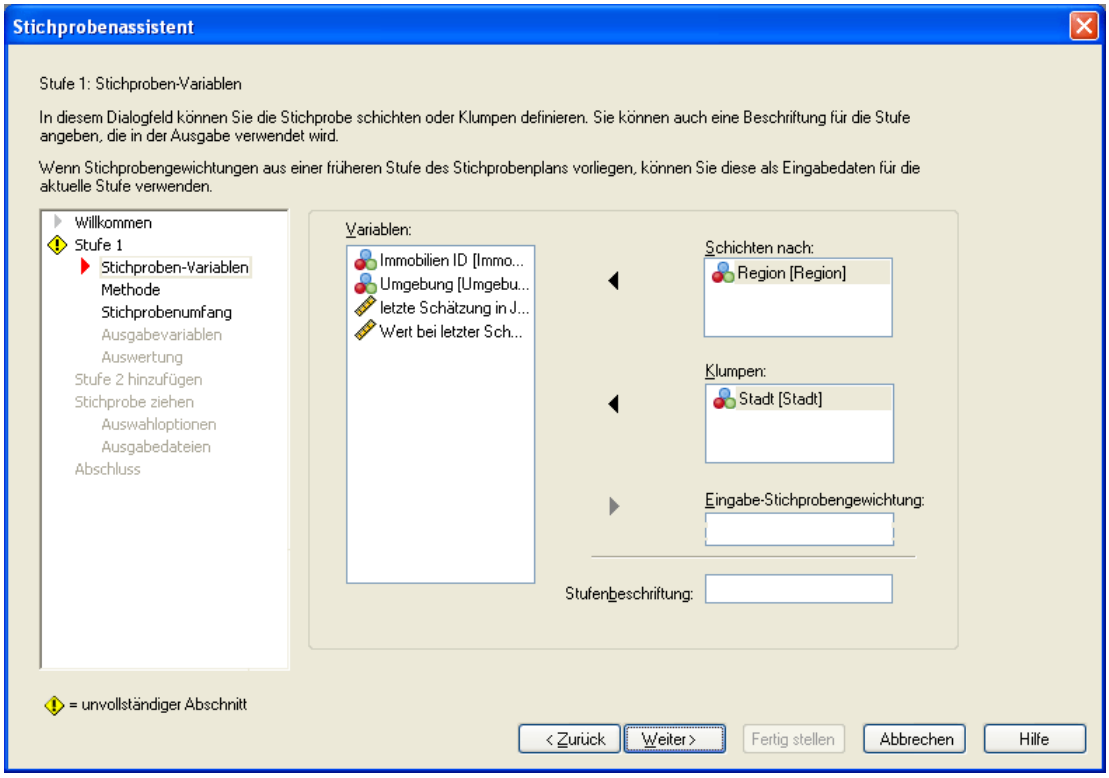
Analysieren
  Komplexe Stichproben
    Stichprobe auswählen...

Abbildung 13-1
Stichprobenassistent – Schritt "Willkommen"


Stichprobenassistent
Willkommen beim Stichprobenassistenten
Der Stichprobenassistent unterstützt Sie bei Entwurf und Auswahl einer komplexen Stichprobe. Ihre Auswahlentscheidungen werden in einer Plandatei gespeichert, die Sie während der Analyse verwenden können, um anzugeben, wie die Datenstichprobe gezogen wurde.
Mit dem Assistenten können Sie außerdem einen Stichprobenplan bearbeiten oder eine Stichprobe anhand eines bestehenden Plans ziehen.
Wie möchten Sie vorgehen?
Stichprobe entwerfen
Wählen Sie diese Option, wenn Sie keine Plandatei erstellt haben. Sie haben dann die Möglichkeit,
Datei:
/property_assess.csplan
Durchsuchen...
Stichprobenplan bearbeiten
Wählen Sie diese Option, wenn Sie Stufen eines bestehenden Plans hinzufügen, entfernen oder bearbeiten möchten. Sie haben
Datei:
Durchsuchen...
Stichprobe ziehen
Wählen Sie diese Option, wenn Sie bereits über eine Plandatei verfügen und eine Stichprobe
Datei:
Durchsuchen...
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

- Wählen Sie die Option Stichprobe entwerfen und geben Sie als Namen der Plandatei /property_assess.csplan ein.
- Klicken Sie auf Weiter.

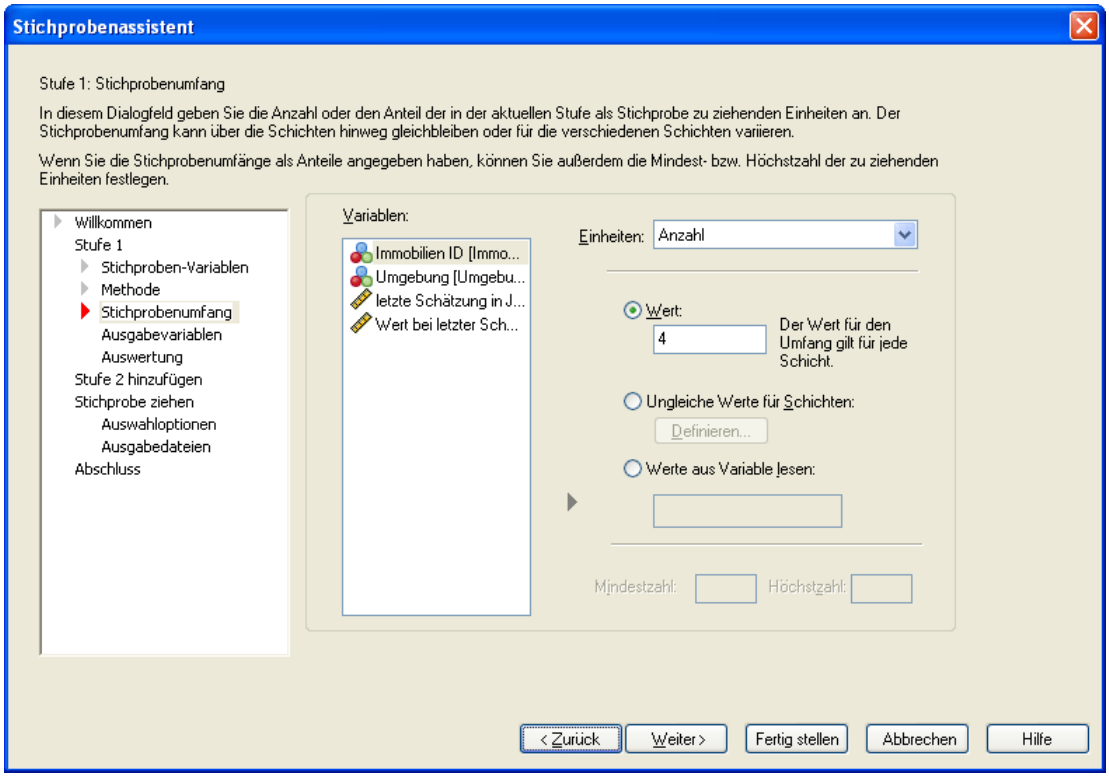
Abbildung 13-2
Stichprobenassistent – Schritt "Stichproben-Variablen" (Stufe 1)



- Wählen Sie *County* als Schichtungsvariable aus.
- Wählen Sie *Township* (Gemeinde) als Klumpenvariable aus.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Methode der Stichprobenziehung" ebenfalls auf Weiter.

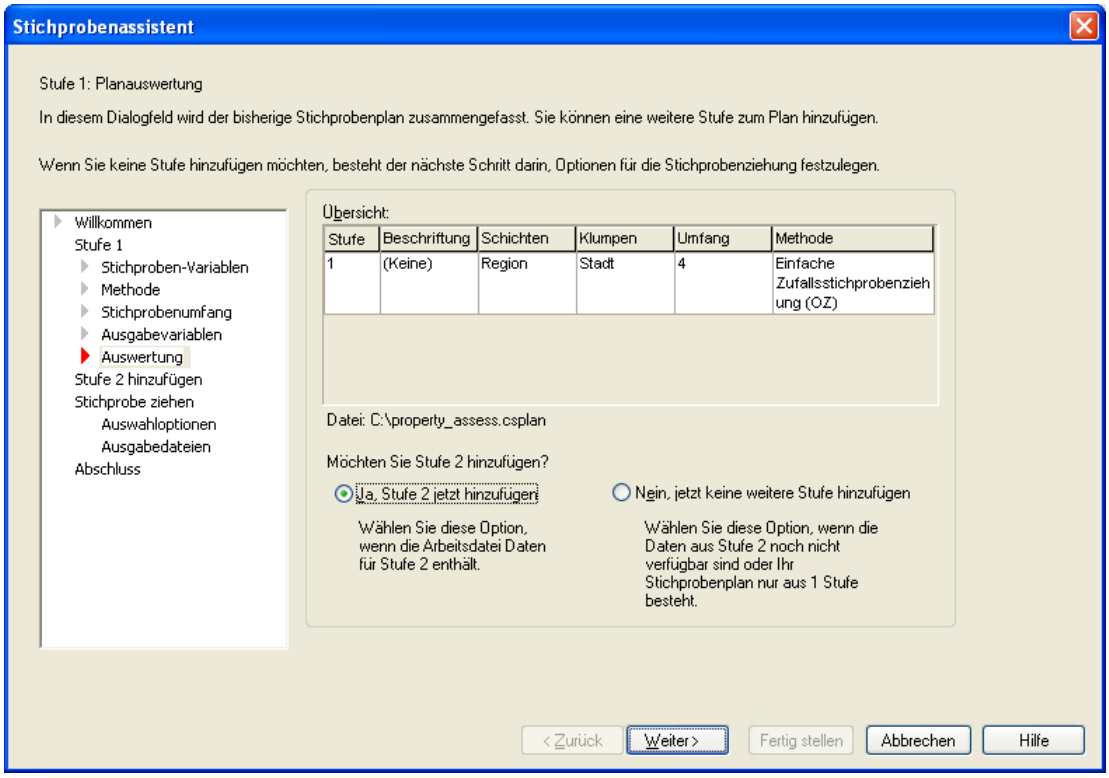
  Diese Planstruktur bedeutet, dass für jedes County unabhängige Stichproben gezogen werden. In dieser Phase werden Gemeinden mithilfe der Standardmethode "Einfache Zufallsstichprobenziehung" als primäre Stichprobeneinheit gezogen.

Abbildung 13-3
Stichprobenassistent, Schritt "Stichprobenumfang" (Stufe 1)



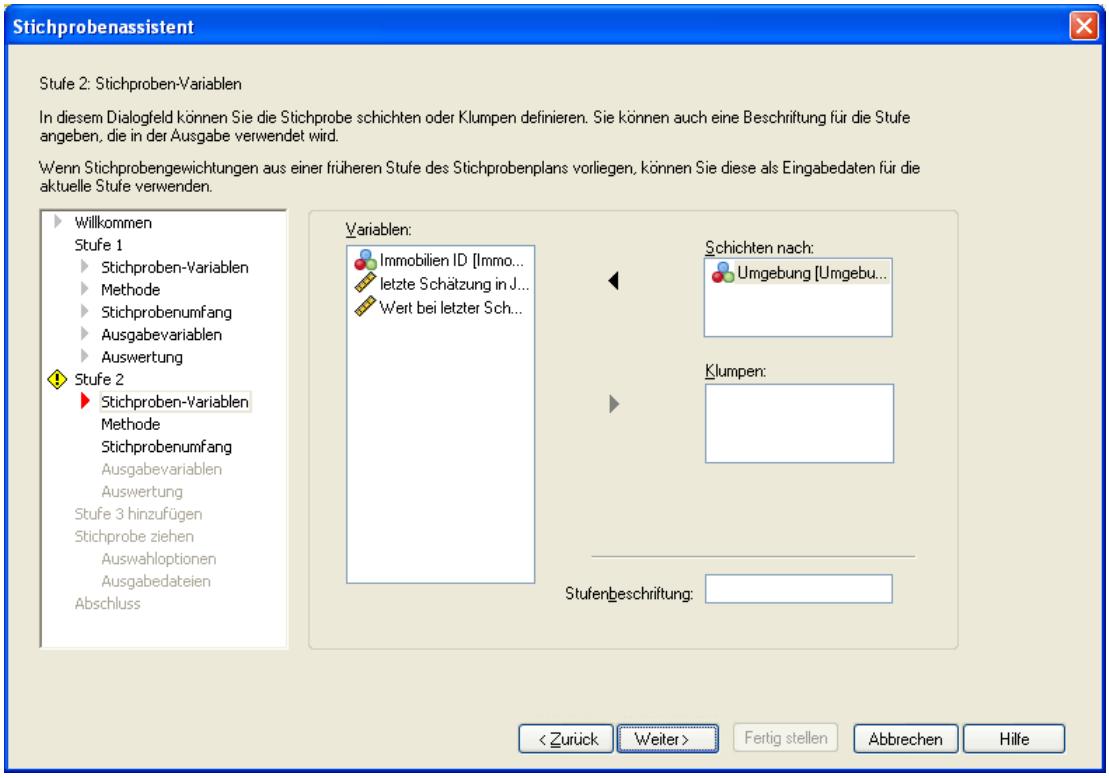
- Wählen Sie in der Dropdown-Liste für die Einheiten Anzahl aus.
- Geben Sie 4 als Wert für die Anzahl der in dieser Stufe auszuwählenden Einheiten ein.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Ausgabevariablen" ebenfalls auf Weiter.

Abbildung 13-4
Stichprobenassistent – Schritt "Planübersicht" (Stufe 1)



- Wählen Sie Ja, Stufe 2 jetzt hinzufügen.
- Klicken Sie auf Weiter.

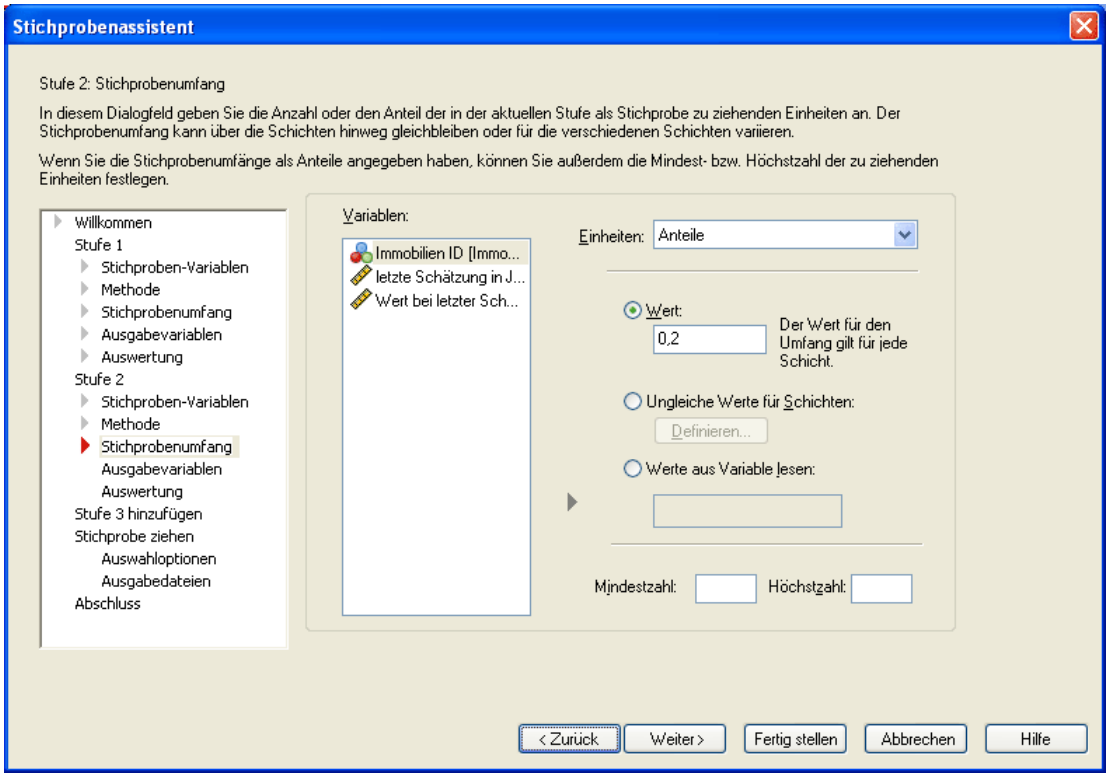
Abbildung 13-5
Stichprobenassistent – Schritt "Stichproben-Variablen" (Stufe 2)



- Wählen Sie *Neighborhood* (Wohnviertel) als Schichtungsvariable aus.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Methode der Stichprobenziehung" ebenfalls auf Weiter.

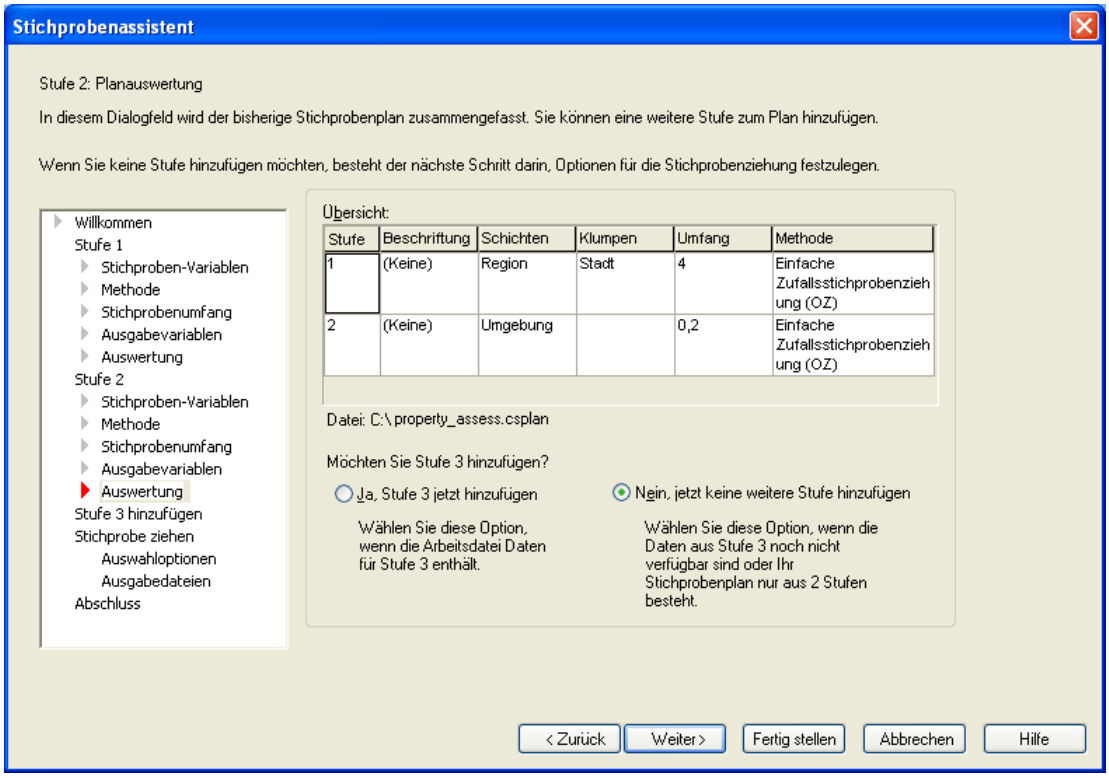
  Diese Planstruktur bedeutet, dass für jedes Wohnviertel der in Stufe 1 ausgewählten Gemeinden unabhängige Stichproben gezogen werden. In dieser Phase werden Immobilien mithilfe der Methode "Einfache Zufallsstichprobenziehung" als primäre Stichprobeneinheit gezogen.

Abbildung 13-6
Stichprobenassistent, Schritt "Stichprobenumfang" (Stufe 2)



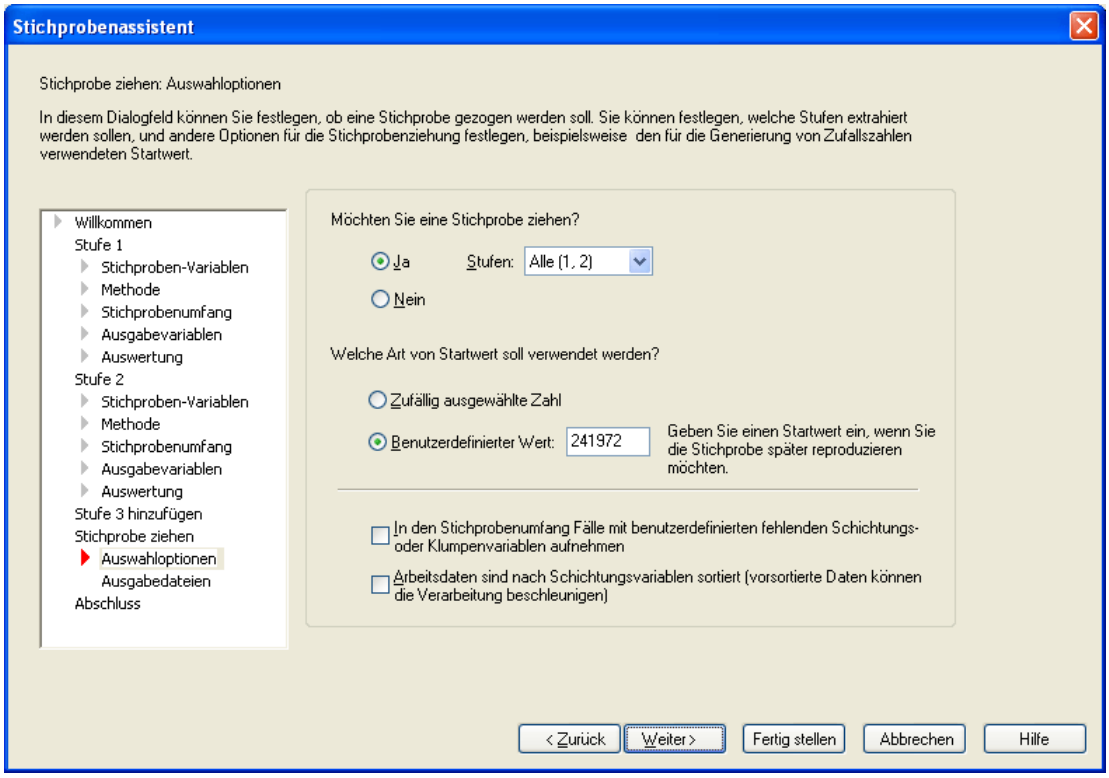
- Wählen Sie in der Dropdown-Liste für die Einheiten Anteile aus.
- Geben Sie 0,2 als Wert für den Anteil der aus jeder Schicht zu ziehenden Einheiten ein.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Ausgabevariablen" ebenfalls auf Weiter.

Abbildung 13-7
Stichprobenassistent – Schritt "Planübersicht" (Stufe 2)



► Überprüfen Sie den Stichprobenplan und klicken Sie dann auf Weiter.

Abbildung 13-8
Stichprobenassistent – Schritt "Stichprobe ziehen: Auswahloptionen"

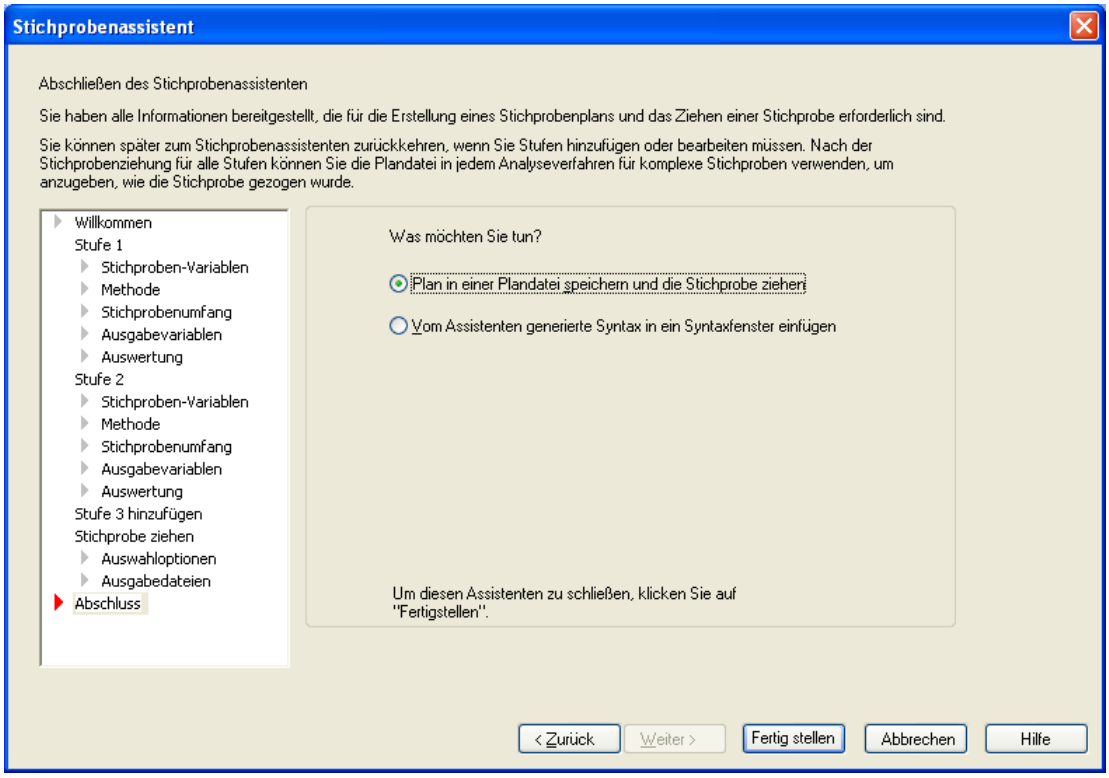


- Wählen Sie Benutzerdefinierter Wert als Typ für den Startwert und geben Sie 241972 als Wert ein.

  Durch die Verwendung eines benutzerdefinierten Werts können Sie die Ergebnisse in diesem Beispiel genau reproduzieren.

- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Stichprobe ziehen: Ausgabedateien" ebenfalls auf Weiter.

Abbildung 13-9
Stichprobenassistent – Schritt "Fertig stellen"


Stichprobenassistent
Abschließen des Stichprobenassistenten
Sie haben alle Informationen bereitgestellt, die für die Erstellung eines Stichprobenplans und das Ziehen einer Stichprobe erforderlich sind.
Sie können später zum Stichprobenassistenten zurückkehren, wenn Sie Stufen hinzufügen oder bearbeiten müssen. Nach der Stichprobenziehung für alle Stufen können Sie die Plandatei in jedem Analyseverfahren für komplexe Stichproben verwenden, um anzugeben, wie die Stichprobe gezogen wurde.
Willkommen
Stufe 1
Stichproben-Variablen
Methode
Stichprobenumfang
Ausgabevariablen
Auswertung
Stufe 2
Stichproben-Variablen
Methode
Stichprobenumfang
Ausgabevariablen
Auswertung
Stufe 3 hinzufügen
Stichprobe ziehen
Auswahloptionen
Ausgabedateien
Abschluss
Was möchten Sie tun?
Plan in einer Plandatei speichern und die Stichprobe ziehen
Vom Assistenten generierte Syntax in ein Syntaxfenster einfügen
Um diesen Assistenten zu schließen, klicken Sie auf "Fertigstellen".
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

▶ Klicken Sie auf Fertig stellen.

Durch diese Auswahl wird die Stichprobenplan-Datei *property_assess.csplan* erstellt und anhand dieses Plans eine Stichprobe gezogen.

Planübersicht

Abbildung 13-10
Planübersicht

			Stufe 1	Stufe 2
Stichproben-Variablen	Schichtung	1	Region	Umgebung
	Klumpen	1	Stadt	
Informationen zur Stichprobe	Auswahlverfahren		Einfache Zufallsstichproben ziehung ohne Zurücklegen	Einfache Zufallsstichproben ziehung ohne Zurücklegen
	Umfang der Stichprobe		4	
	Erzeugte oder veränderte Variablen	Stufenweise Einschluss-(Auswahl-)wahrscheinlichkeiten	Inclusion Probability_1_	Inclusion Probability_2_
		Stufenweise kumulierte Stichprobengewichtung	Sample Weight Cumulative_1_	Sample Weight Cumulative_2_
	Anteil der Stichprobe			,2
Informationen für die Auswertung	Annahmen für die Schätzung		Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen	Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen
	Einschlußwahrscheinlichkeit		Gewonnen aus Variable Inclusion Probability_1_	Gewonnen aus Variable Inclusion Probability_2_

Plandatei: C:\property_assess.csplan
Gewichtungsvariable: SampleWeight_Final_

Die zusammenfassende Tabelle enthält eine Übersicht über den Stichprobenplan. Anhand dieser Tabelle können Sie überprüfen, ob der Plan tatsächlich Ihren Absichten entspricht.

Stichprobenübersicht

Abbildung 13-11
Stufenübersicht

	Umfang der Stichprobe		Anteil der Stichprobe	
Region	Gewünscht	Tatsächlich	Gewünscht	Tatsächlich
Osten	4	4	44,4%	44,4%
Mitte	4	4	57,1%	57,1%
Westen	4	4	25,0%	25,0%
Norden	4	4	44,4%	44,4%
Süden	4	4	50,0%	50,0%

Plandatei: C:\property_assess.csplan

Diese zusammenfassende Tabelle enthält eine Übersicht über die erste Stufe der Stichprobenziehung. Anhand dieser Tabelle können Sie überprüfen, ob die Stichprobenziehung plangemäß verlief. Wie angefordert wurden aus jedem County vier Gemeinden als Stichprobe gezogen.

Abbildung 13-12
Stufenübersicht

			Umfang der Stichprobe		Anteil der Stichprobe	
Region	Stadt	Umgebung	Gewünscht	Tatsächlich	Gewünscht	Tatsächlich
Osten	2	8	4	4	20,0%	19,0%
		9	14	14	20,0%	20,6%
		10	7	7	20,0%	18,9%
		11	14	14	20,0%	20,0%
	6	36	13	13	20,0%	20,3%
		37	14	14	20,0%	20,6%
		38	13	13	20,0%	20,6%
	7	43	12	12	20,0%	20,7%
		44	11	11	20,0%	19,6%
		45	11	11	20,0%	20,8%
		46	13	13	20,0%	20,0%
	9	57	13	13	20,0%	20,6%
		58	5	5	20,0%	18,5%
		59	11	11	20,0%	19,3%
		60	13	13	20,0%	19,4%
Mitte	22	148	9	9	20,0%	19,6%
		149	8	8	20,0%	20,0%
		150	14	14	20,0%	20,0%
		151	11	11	20,0%	19,6%

Diese zusammenfassende Tabelle (deren oberer Teil hier zu sehen ist) enthält eine Übersicht über die zweite Stufe der Stichprobenziehung. Außerdem können Sie damit überprüfen, ob die Stichprobenziehung plangemäß verlief. Wie angefordert wurden ungefähr 20 % der Immobilien aus jedem Wohnviertel in jeder der in der ersten Stufe ausgewählten Gemeinden als Stichprobe gezogen.

Stichprobenergebnisse

Abbildung 13-13
Daten-Editor mit Stichprobenergebnissen

	propid	nbrhood	town	county	time	lastval	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
273	577,0	8	2	1	4	181,70	.	.	.	.	.
274	578,0	8	2	1	5	189,60	.	.	.	.	.
275	579,0	8	2	1	4	200,10	.	.	.	.	.
276	580,0	8	2	1	5	211,50	.	.	.	.	.
277	581,0	8	2	1	4	181,50	.	.	.	.	.
278	641,0	9	2	1	7	192,40	.	.	.	.	.
279	642,0	9	2	1	6	236,70	,44	2,25	,21	10,93	10,93
280	643,0	9	2	1	6	150,40	,44	2,25	,21	10,93	10,93
281	644,0	9	2	1	8	204,80	.	.	.	.	.
282	645,0	9	2	1	6	225,40	.	.	.	.	.
283	646,0	9	2	1	7	180,80	,44	2,25	,21	10,93	10,93

Datenansicht | Variablenansicht

Die Ergebnisse der Stichprobenziehung werden im Daten-Editor angezeigt. Fünf neue Variablen wurden in der Arbeitsdatei gespeichert. Diese stehen für die Einschlusswahrscheinlichkeiten und die kumulierten Stichprobengewichtungen für die einzelnen Stufen sowie für die endgültigen Stichprobengewichtungen.

- Fälle mit Werten für diese Variablen wurden für die Stichprobe ausgewählt.
- Fälle mit systemdefinierten fehlenden Werten für die Variablen wurden nicht ausgewählt.

Die Behörde verwendet nun ihre Ressourcen, um aktuelle Bewertungen für die in der Stichprobe ausgewählten Immobilien einzuholen. Sobald diese Bewertungen vorliegen, kann die Stichprobe mit den Analyseverfahren für komplexe Stichproben verarbeitet werden. Die Stichprobenspezifikationen entnehmen Sie dem Stichprobenplan *property_assess.csplan*.

Ziehen einer Stichprobe aus einem partiellen Stichprobenrahmen

Ein Unternehmen ist daran interessiert, eine Datenbank mit qualitativ hochwertigen Umfrageinformationen zusammenzustellen und zu verkaufen. Die Umfragestichprobe sollte repräsentativ sein, aber die Stichprobenziehung soll dennoch effizient sein. Daher werden Methoden für komplexe Stichproben verwendet. Ein vollständiger Stichprobenplan würde zu folgender Struktur führen:

Stufe	Schichten	Klumpen
1	Region	Provinz
2	Bezirk	Ort
3	Wohngebiet	

In der dritten Stufe sind Haushalte die primäre Stichprobeneinheit und die Umfrage wird in den ausgewählten Haushalten durchgeführt. Da Informationen jedoch nur auf der Ortsebene leicht zu beschaffen sind, hat das Unternehmen vor, die ersten beiden Stufen des Plans jetzt durchzuführen und anschließend Informationen zur Anzahl der Wohngebiete und Haushalte in den als Stichprobe ausgewählten Orten einzuholen. Die auf der Ortsebene zur Verfügung stehenden Informationen befinden sich in der Datei *demo_cs_1.sav*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Beachten Sie, dass diese Datei eine Variable *Wohngebiet* enthält, die überall den Wert 1 aufweist. Hierbei handelt es sich um einen Platzhalter für die Variable “wahr”, deren Werte nach der Ausführung der ersten beiden Stufen des Stichprobenplans erfasst werden. Mit diesem Platzhalter können Sie bereits jetzt den vollständigen, drei Stufen umfassenden Stichprobenplan angeben. Geben Sie mithilfe des Stichprobenassistenten für komplexe Stichproben den vollständigen komplexen Stichprobenplan an und ziehen Sie anschließend die Stichproben für die ersten beiden Stufen.

Verwenden des Assistenten für die Stichprobenziehung aus dem ersten Teilrahmen

▶ Um den Stichprobenassistenten für komplexe Stichproben durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

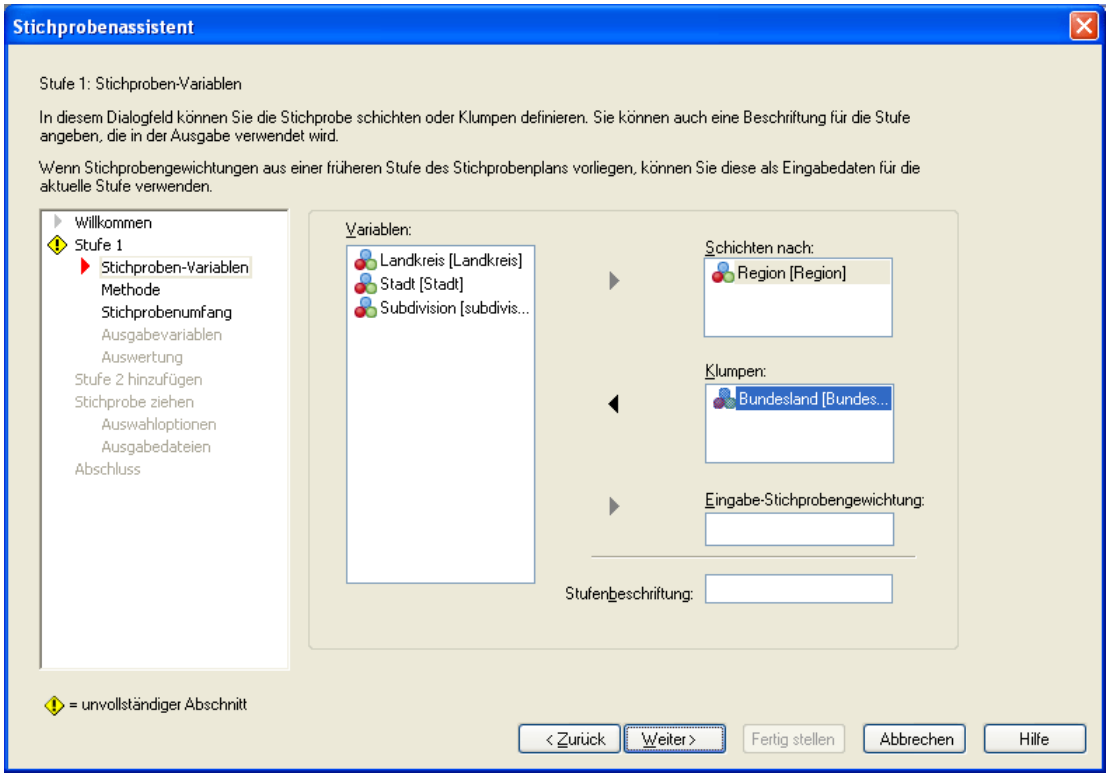
Analysieren
  Komplexe Stichproben
    Stichprobe auswählen...

Abbildung 13-14
Stichprobenassistent – Schritt "Willkommen"

Stichprobenassistent
Willkommen beim Stichprobenassistenten
Der Stichprobenassistent unterstützt Sie bei Entwurf und Auswahl einer komplexen Stichprobe. Ihre Auswahlentscheidungen werden in einer Plandatei gespeichert, die Sie während der Analyse verwenden können, um anzugeben, wie die Datenstichprobe gezogen wurde.
Mit dem Assistenten können Sie außerdem einen Stichprobenplan bearbeiten oder eine Stichprobe anhand eines bestehenden Plans ziehen.
Wie möchten Sie vorgehen?
Stichprobe entwerfen
Wählen Sie diese Option, wenn Sie keine Plandatei erstellt haben. Sie haben dann die Möglichkeit,
Datei:
\demo.csplan
Durchsuchen...
Stichprobenplan bearbeiten
Wählen Sie diese Option, wenn Sie Stufen eines bestehenden Plans hinzufügen, entfernen oder bearbeiten möchten. Sie haben
Datei:
Durchsuchen...
Stichprobe ziehen
Wählen Sie diese Option, wenn Sie bereits über eine Plandatei verfügen und eine Stichprobe
Datei:
Durchsuchen...
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

- Wählen Sie die Option Stichprobe entwerfen und geben Sie als Namen der Plandatei /demo.csplan ein.
- Klicken Sie auf Weiter.

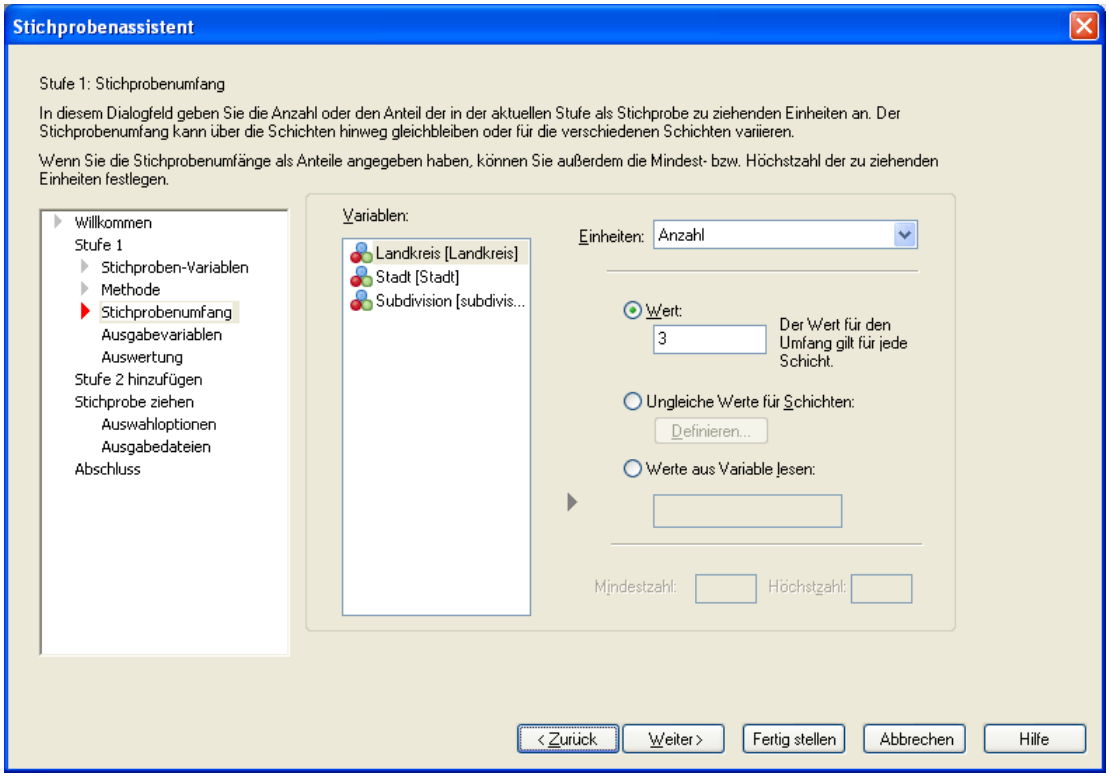
Abbildung 13-15
Stichprobenassistent – Schritt "Stichproben-Variablen" (Stufe 1)



- Wählen Sie *Region* als Schichtungsvariable aus.
- Wählen Sie *Province* (Provinz) als Klumpenvariable aus.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Methode der Stichprobenziehung" ebenfalls auf Weiter.

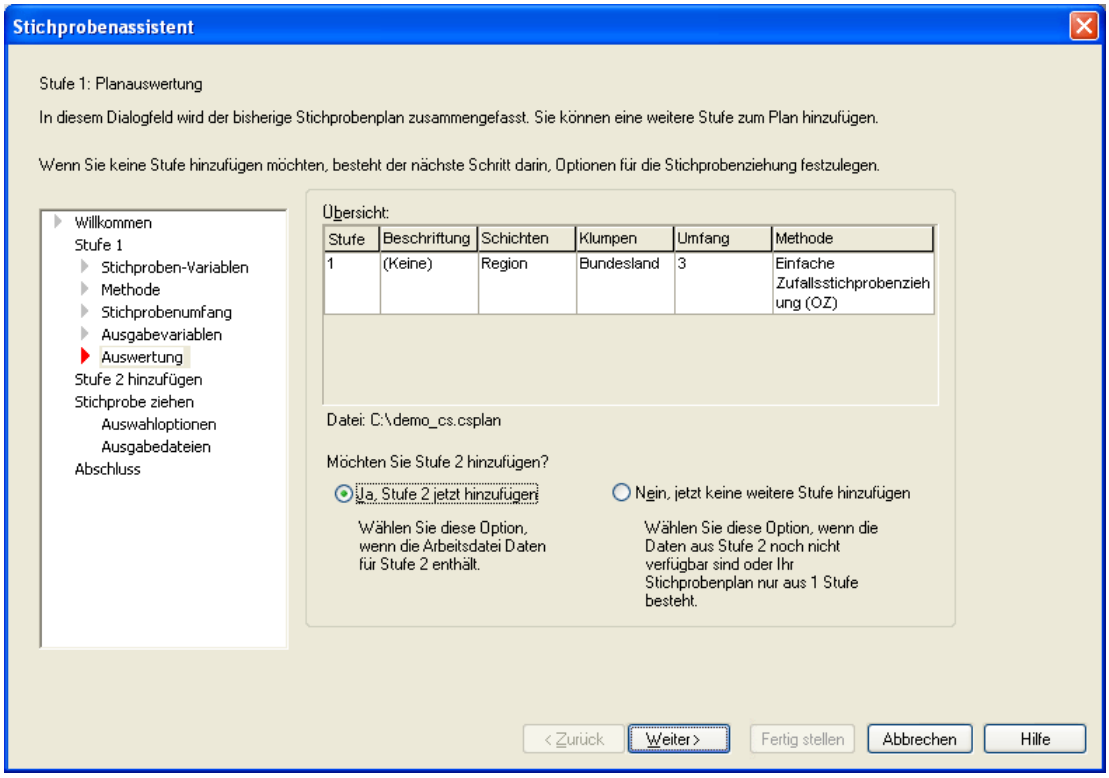
  Diese Planstruktur bedeutet, dass für jede Region unabhängige Stichproben gezogen werden. In dieser Phase werden Provinzen mithilfe der Standardmethode "Einfache Zufallsstichprobenziehung" als primäre Stichprobeneinheit gezogen.

Abbildung 13-16
Stichprobenassistent, Schritt "Stichprobenumfang" (Stufe 1)



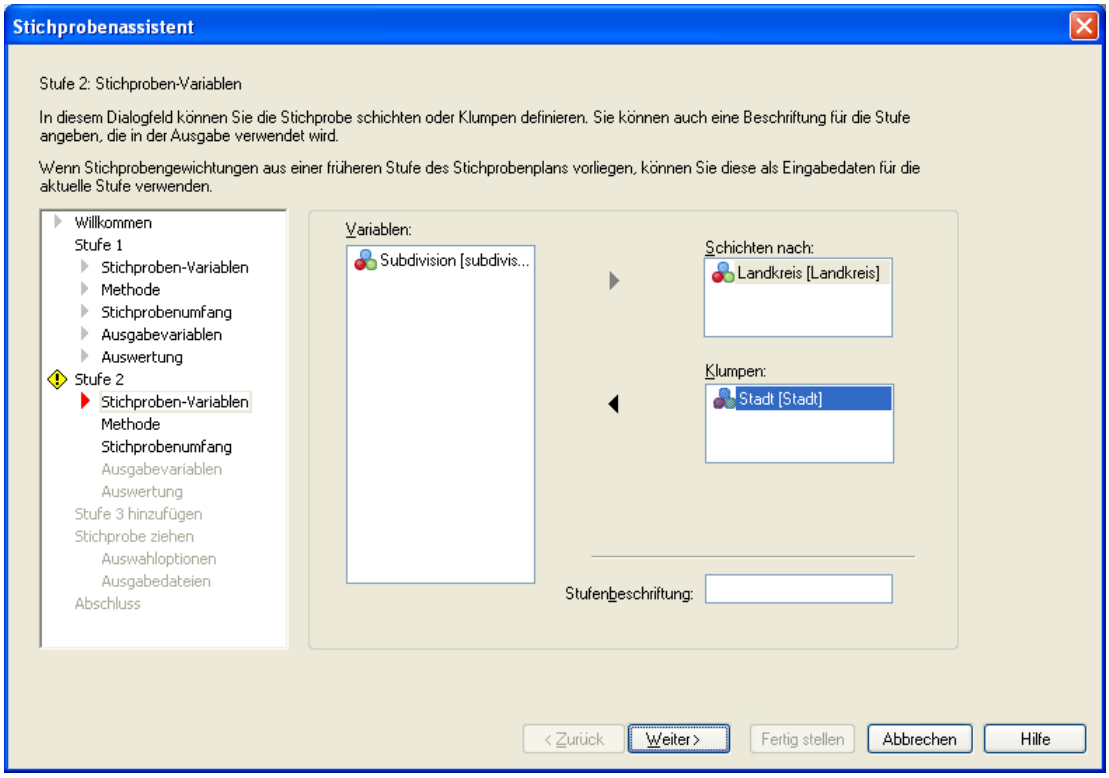
- Wählen Sie in der Dropdown-Liste für die Einheiten Anzahl aus.
- Geben Sie 3 als Wert für die Anzahl der in dieser Stufe auszuwählenden Einheiten ein.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Ausgabevariablen" ebenfalls auf Weiter.

Abbildung 13-17
Stichprobenassistent – Schritt "Planübersicht" (Stufe 1)



- Wählen Sie Ja, Stufe 2 jetzt hinzufügen.
- Klicken Sie auf Weiter.

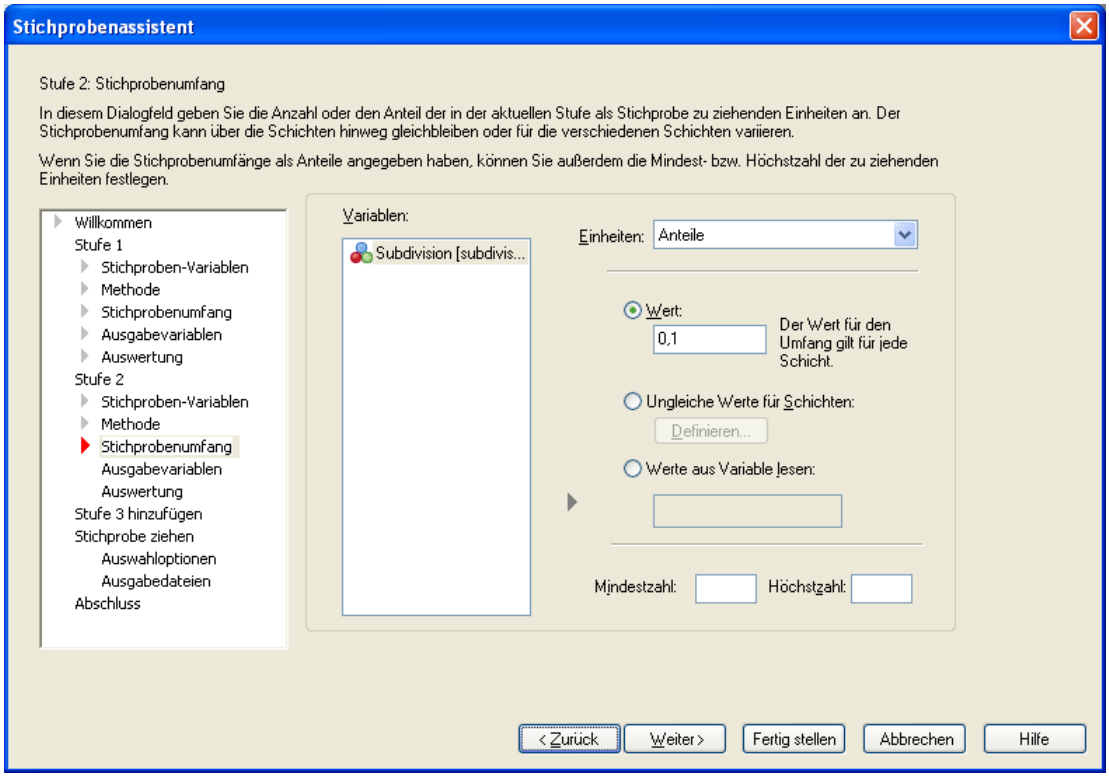
Abbildung 13-18
Stichprobenassistent – Schritt "Stichproben-Variablen" (Stufe 2)



- Wählen Sie *District* (Bezirk) als Schichtungsvariable aus.
- Wählen Sie *City* (Ort) als Klumpenvariable aus.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Methode der Stichprobenziehung" ebenfalls auf Weiter.

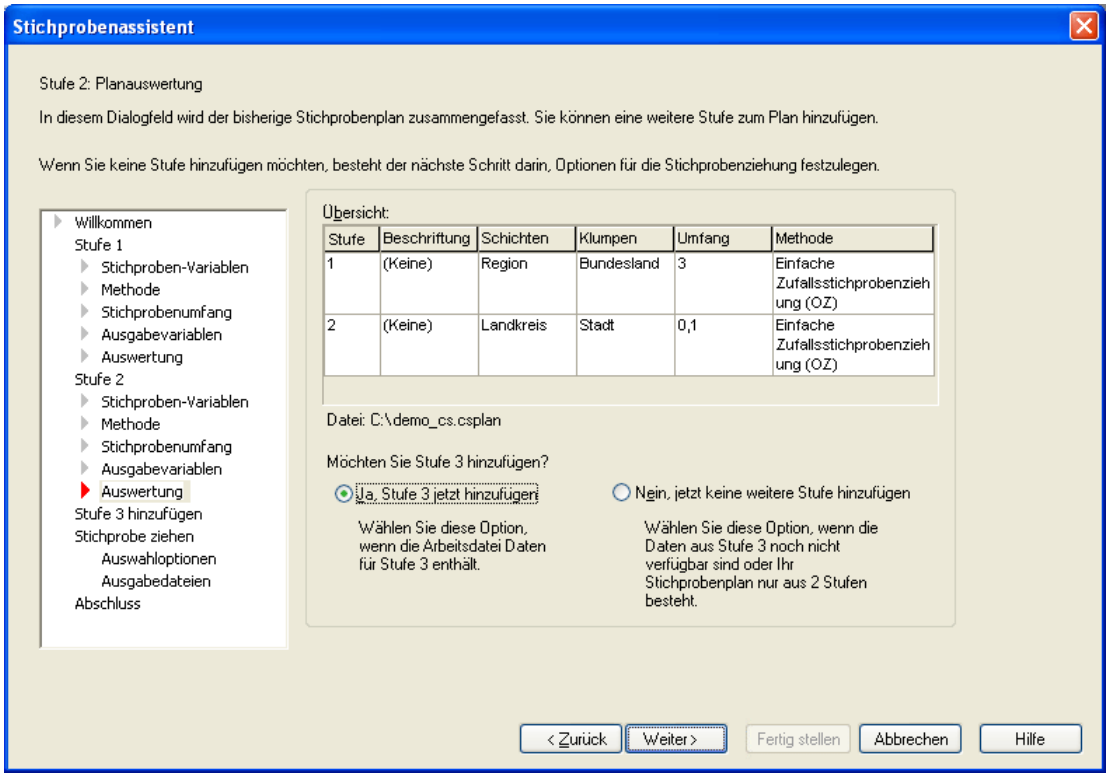
  Diese Planstruktur bedeutet, dass für jeden Bezirk unabhängige Stichproben gezogen werden. In dieser Phase werden Orte mithilfe der Standardmethode "Einfache Zufallsstichprobenziehung" als primäre Stichprobeneinheit gezogen.

Abbildung 13-19
Stichprobenassistent, Schritt "Stichprobenumfang" (Stufe 2)



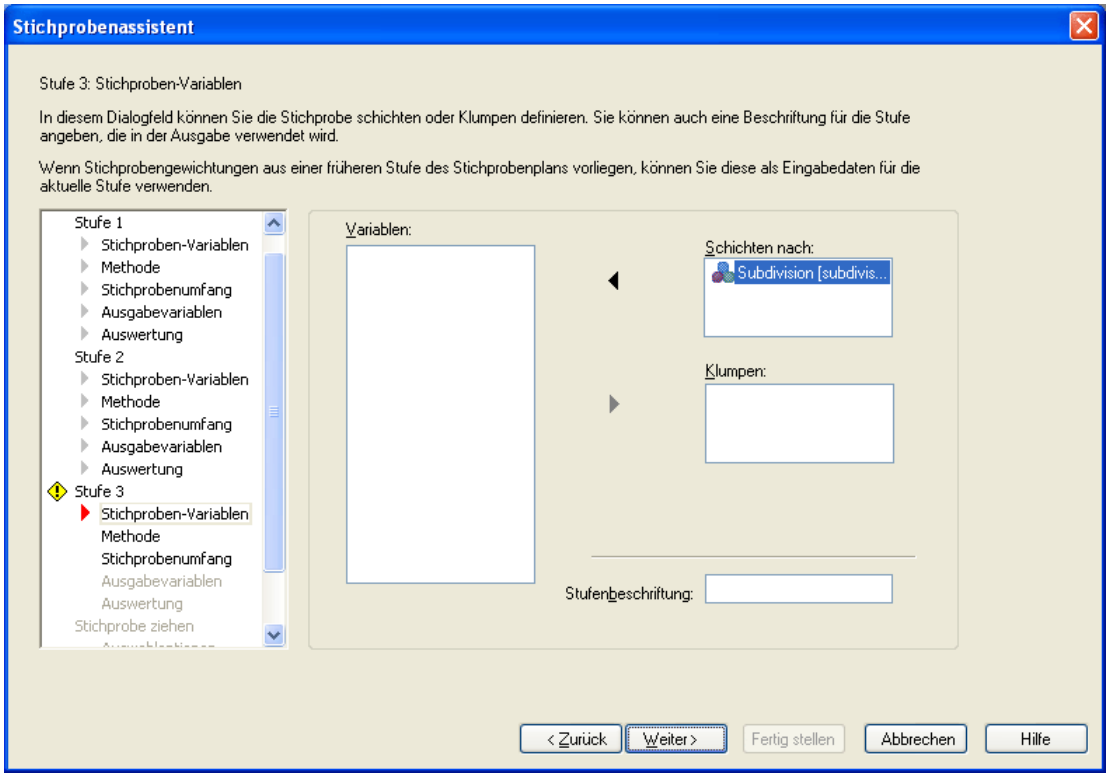
- Wählen Sie in der Dropdown-Liste für die Einheiten Anteile aus.
- Geben Sie 0,1 als Wert für den Anteil der aus jeder Schicht zu ziehenden Einheiten ein.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Ausgabevariablen" ebenfalls auf Weiter.

Abbildung 13-20
Stichprobenassistent – Schritt "Planübersicht" (Stufe 2)


Stichprobenassistent
Stufe 2: Planauswertung
In diesem Dialogfeld wird der bisherige Stichprobenplan zusammengefasst. Sie können eine weitere Stufe zum Plan hinzufügen.
Wenn Sie keine Stufe hinzufügen möchten, besteht der nächste Schritt darin, Optionen für die Stichprobenziehung festzulegen.
Willkommen
Stufe 1
Stichproben-Variablen
Methode
Stichprobenumfang
Ausgabevariablen
Auswertung
Stufe 2
Stichproben-Variablen
Methode
Stichprobenumfang
Ausgabevariablen
Auswertung
Stufe 3 hinzufügen
Stichprobe ziehen
Auswahloptionen
Ausgabedateien
Abschluss
Übersicht:
Stufe | Beschriftung | Schichten | Klumpen | Umfang | Methode
1 | (Keine) | Region | Bundesland | 3 | Einfache Zufallsstichprobenziehung (OZ)
2 | (Keine) | Landkreis | Stadt | 0,1 | Einfache Zufallsstichprobenziehung (OZ)
Datei: C:\demo_cs.csplan
Möchten Sie Stufe 3 hinzufügen?
Ja, Stufe 3 jetzt hinzufügen
Wählen Sie diese Option, wenn die Arbeitsdatei Daten für Stufe 3 enthält.
Nein, jetzt keine weitere Stufe hinzufügen
Wählen Sie diese Option, wenn die Daten aus Stufe 3 noch nicht verfügbar sind oder Ihr Stichprobenplan nur aus 2 Stufen besteht.
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- Wählen Sie Ja, Stufe 3 jetzt hinzufügen.
- Klicken Sie auf Weiter.

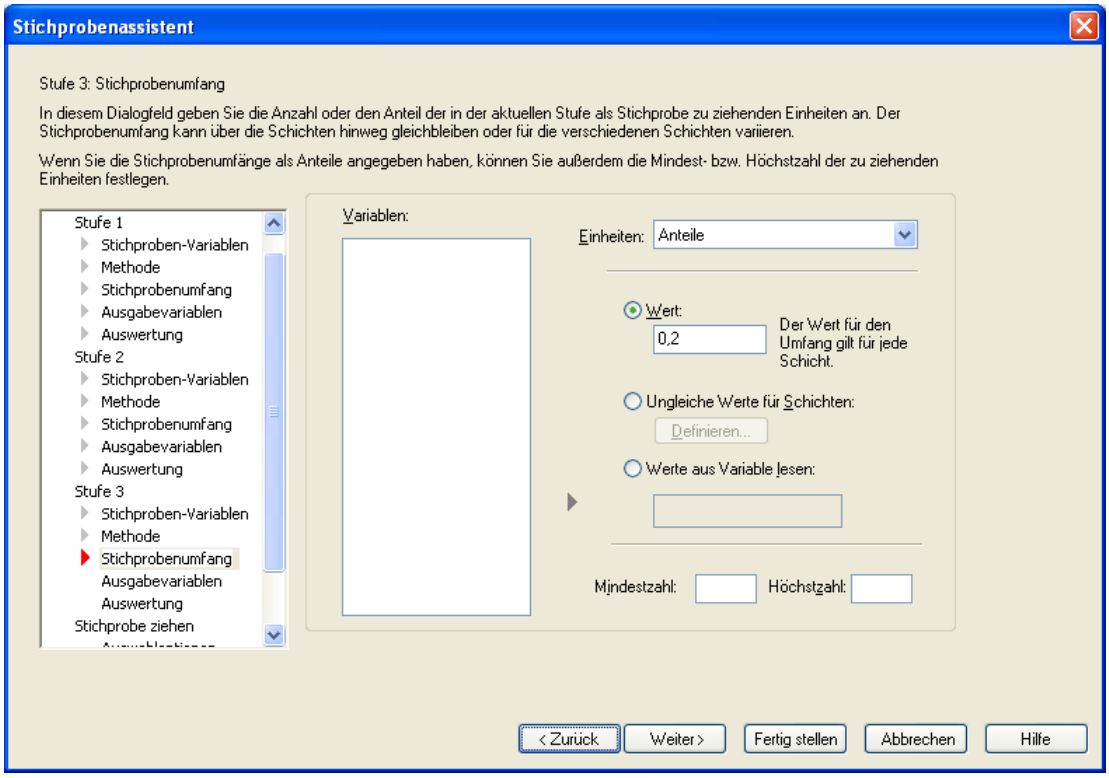
Abbildung 13-21
Stichprobenassistent – Schritt "Stichproben-Variablen" (Stufe 3)



- Wählen Sie *Wohngebiet* als Schichtungsvariable aus.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Methode der Stichprobenziehung" ebenfalls auf Weiter.

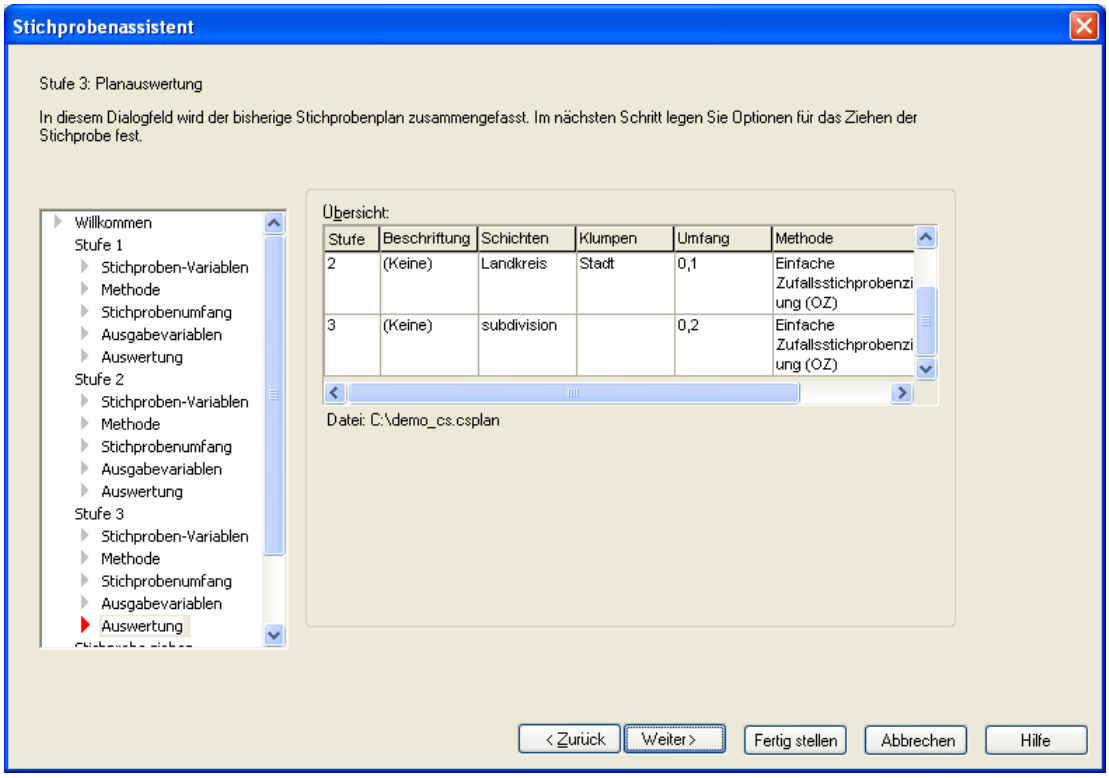
  Diese Planstruktur bedeutet, dass für jedes Wohngebiet unabhängige Stichproben gezogen werden. In dieser Phase werden Haushaltseinheiten mithilfe der Standardmethode "Einfache Zufallsstichprobenziehung" als primäre Stichprobeneinheit gezogen.

Abbildung 13-22
Stichprobenassistent, Schritt "Stichprobenumfang" (Stufe 3)



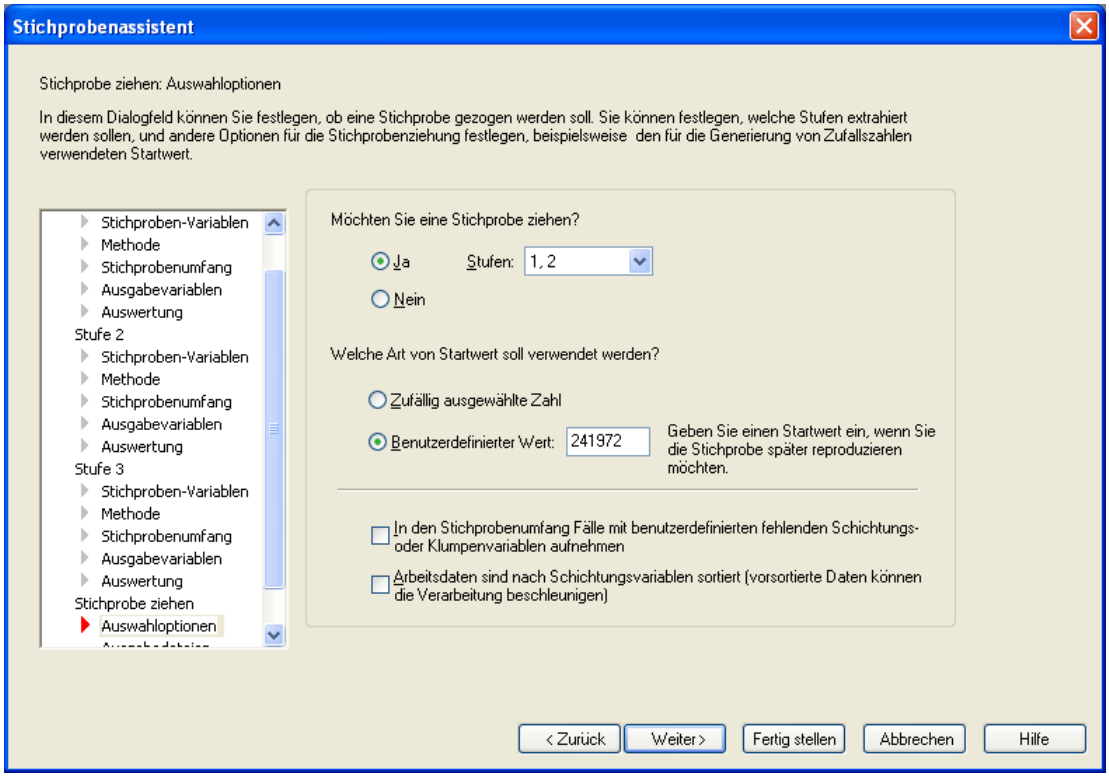
- Wählen Sie in der Dropdown-Liste für die Einheiten Anteile aus.
- Geben Sie 0,2 als Wert für den Anteil der in dieser Stufe auszuwählenden Einheiten ein.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Ausgabevariablen" ebenfalls auf Weiter.

Abbildung 13-23
Stichprobenassistent – Schritt "Planübersicht" (Stufe 3)



- Überprüfen Sie den Stichprobenplan und klicken Sie dann auf **Weiter**.

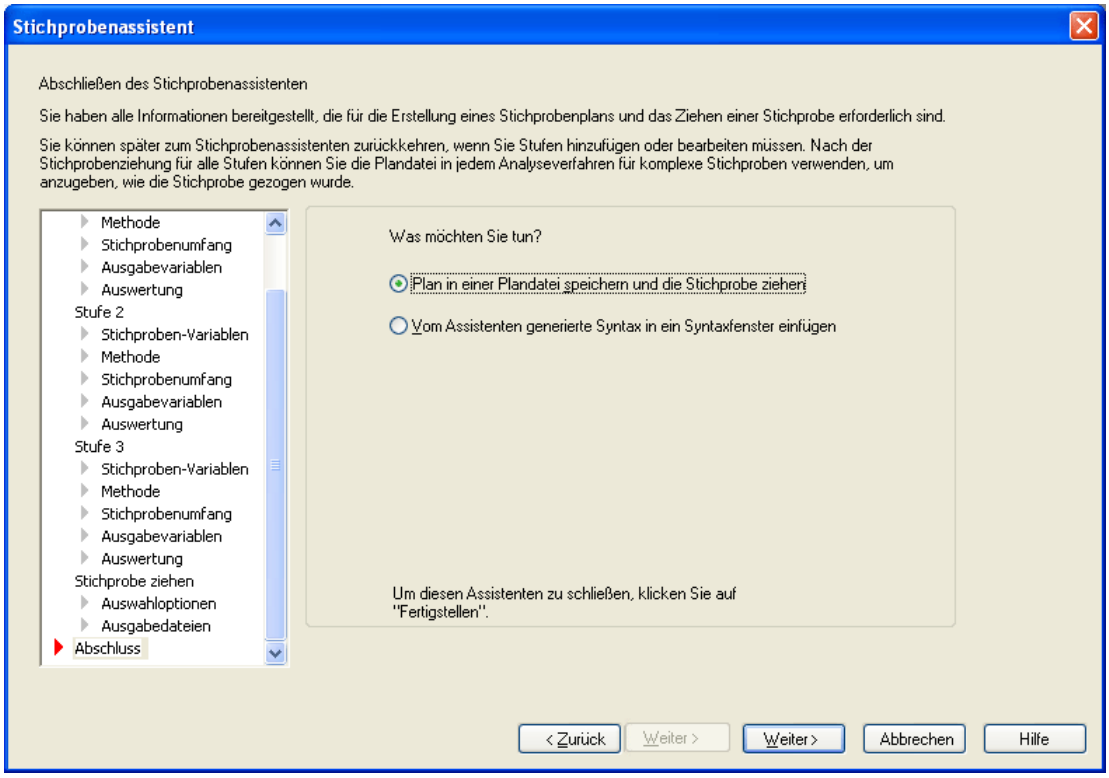
Abbildung 13-24
Stichprobenassistent – Schritt "Stichprobe ziehen: Auswahloptionen"



- Wählen Sie 1, 2 als Stufen, für die nun die Stichproben gezogen werden sollen.
- Wählen Sie Benutzerdefinierter Wert als Typ für den Startwert und geben Sie 241972 als Wert ein.

  Durch die Verwendung eines benutzerdefinierten Werts können Sie die Ergebnisse in diesem Beispiel genau reproduzieren.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Stichprobe ziehen: Ausgabedateien" ebenfalls auf Weiter.

Abbildung 13-25
Stichprobenassistent – Schritt "Fertig stellen"

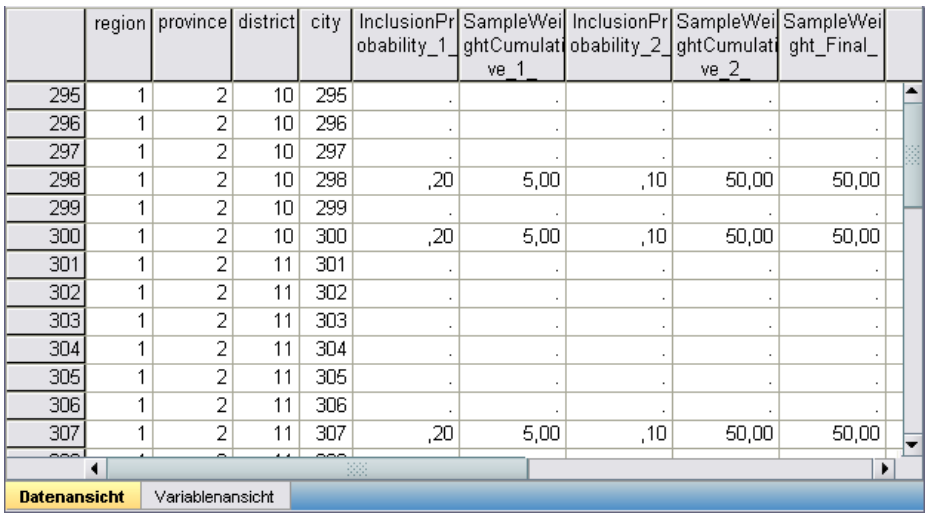


▶ Klicken Sie auf Fertig stellen.

Durch diese Auswahl wird die Stichprobenplan-Datei *demo.csplan* erstellt und anhand der ersten beiden Stufen dieses Plans eine Stichprobe gezogen.

Stichprobenergebnisse

Abbildung 13-26
Daten-Editor mit Stichprobenergebnissen



	region	province	district	city	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
295	1	2	10	295	.	.	.	.	.
296	1	2	10	296	.	.	.	.	.
297	1	2	10	297	.	.	.	.	.
298	1	2	10	298	,20	5,00	,10	50,00	50,00
299	1	2	10	299	.	.	.	.	.
300	1	2	10	300	,20	5,00	,10	50,00	50,00
301	1	2	11	301	.	.	.	.	.
302	1	2	11	302	.	.	.	.	.
303	1	2	11	303	.	.	.	.	.
304	1	2	11	304	.	.	.	.	.
305	1	2	11	305	.	.	.	.	.
306	1	2	11	306	.	.	.	.	.
307	1	2	11	307	,20	5,00	,10	50,00	50,00

Datenansicht | Variablenansicht

Die Ergebnisse der Stichprobenziehung werden im Daten-Editor angezeigt. Fünf neue Variablen wurden in der Arbeitsdatei gespeichert. Diese stehen für die Einschlusswahrscheinlichkeiten und die kumulierten Stichprobengewichtungen für die einzelnen Stufen sowie für die "endgültigen" Stichprobengewichtungen für die ersten beiden Stufen.

- Orte mit Werten für diese Variablen wurden für die Stichprobe ausgewählt.
- Orte mit systemdefinierten fehlenden Werten für die Variablen wurden nicht ausgewählt.

Für jeden ausgewählten Ort hat die Firma Informationen zu den Wohngebieten und den Haushaltseinheiten eingeholt und in der Datei *demo_cs_2.sav* gespeichert. Verwenden Sie diese Datei und den Stichprobenassistenten, um die Stichprobenziehung für die dritte Stufe dieses Plans durchzuführen.

Verwenden des Assistenten für die Stichprobenziehung aus dem zweiten Teilrahmen

▶ Um den Stichprobenassistenten für komplexe Stichproben durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

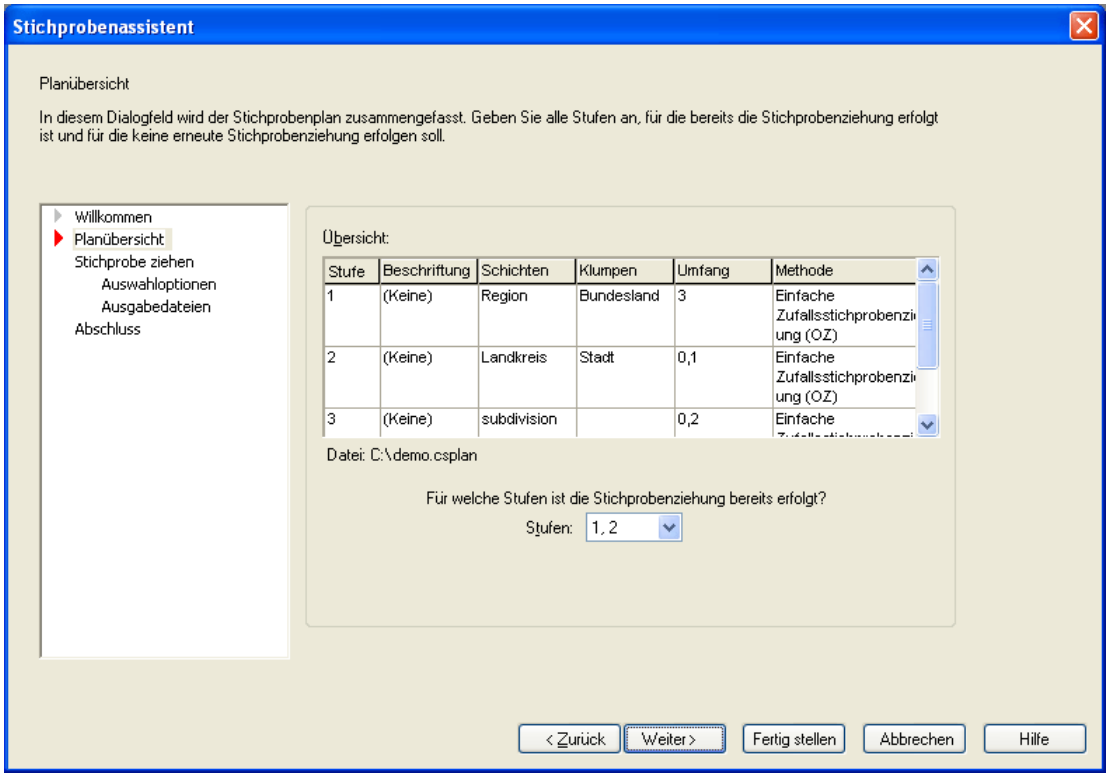
Analysieren
  Komplexe Stichproben
    Stichprobe auswählen...

Abbildung 13-27
Stichprobenassistent – Schritt "Willkommen"

Stichprobenassistent
Willkommen beim Stichprobenassistenten
Der Stichprobenassistent unterstützt Sie bei Entwurf und Auswahl einer komplexen Stichprobe. Ihre Auswahlentscheidungen werden in einer Plandatei gespeichert, die Sie während der Analyse verwenden können, um anzugeben, wie die Datenstichprobe gezogen wurde.
Mit dem Assistenten können Sie außerdem einen Stichprobenplan bearbeiten oder eine Stichprobe anhand eines bestehenden Plans ziehen.
Wie möchten Sie vorgehen?
Stichprobe entwerfen
Wählen Sie diese Option, wenn Sie keine Plandatei erstellt haben. Sie haben dann die Möglichkeit,
Datei:
Durchsuchen...
Stichprobenplan bearbeiten
Wählen Sie diese Option, wenn Sie Stufen eines bestehenden Plans hinzufügen, entfernen oder bearbeiten möchten. Sie haben
Datei:
Durchsuchen...
Stichprobe ziehen
Wählen Sie diese Option, wenn Sie bereits über eine Plandatei verfügen und eine Stichprobe
Datei:
\demo.csplan
Durchsuchen...
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

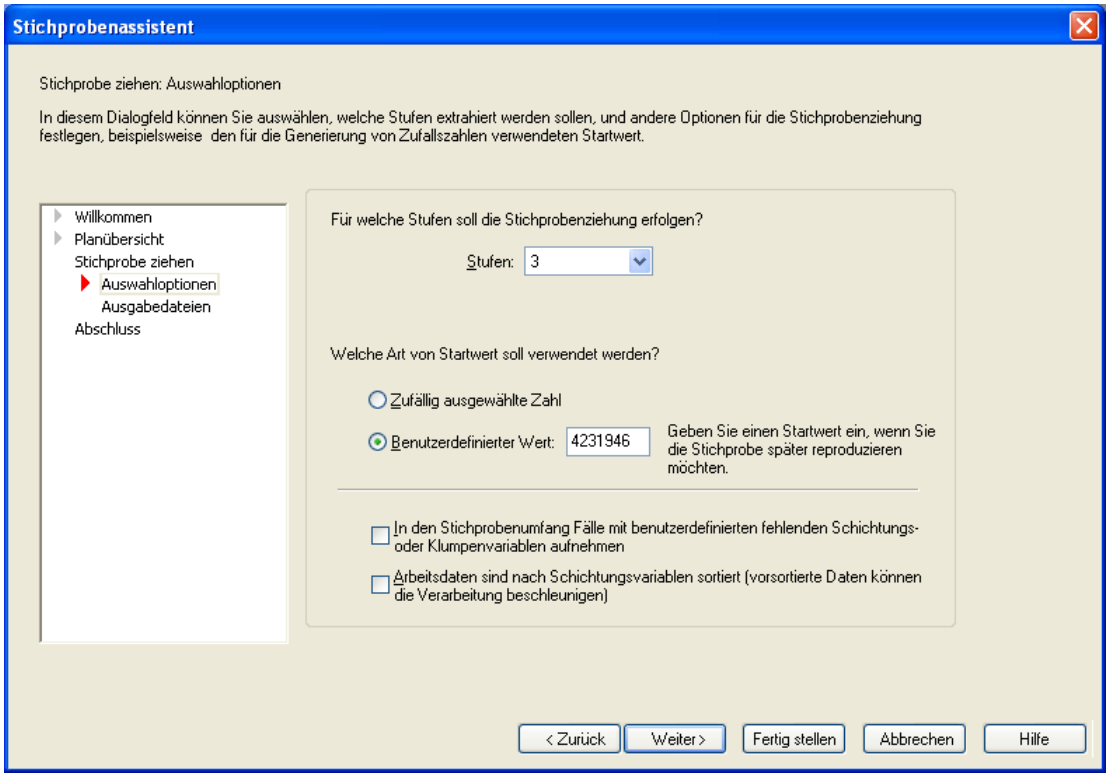
- Wählen Sie die Option Stichprobe ziehen und wählen Sie die von Ihnen erstellte Plandatei /demo.csplan aus.
- Klicken Sie auf Weiter.

Abbildung 13-28
Stichprobenassistent – Schritt "Planübersicht" (Stufe 3)



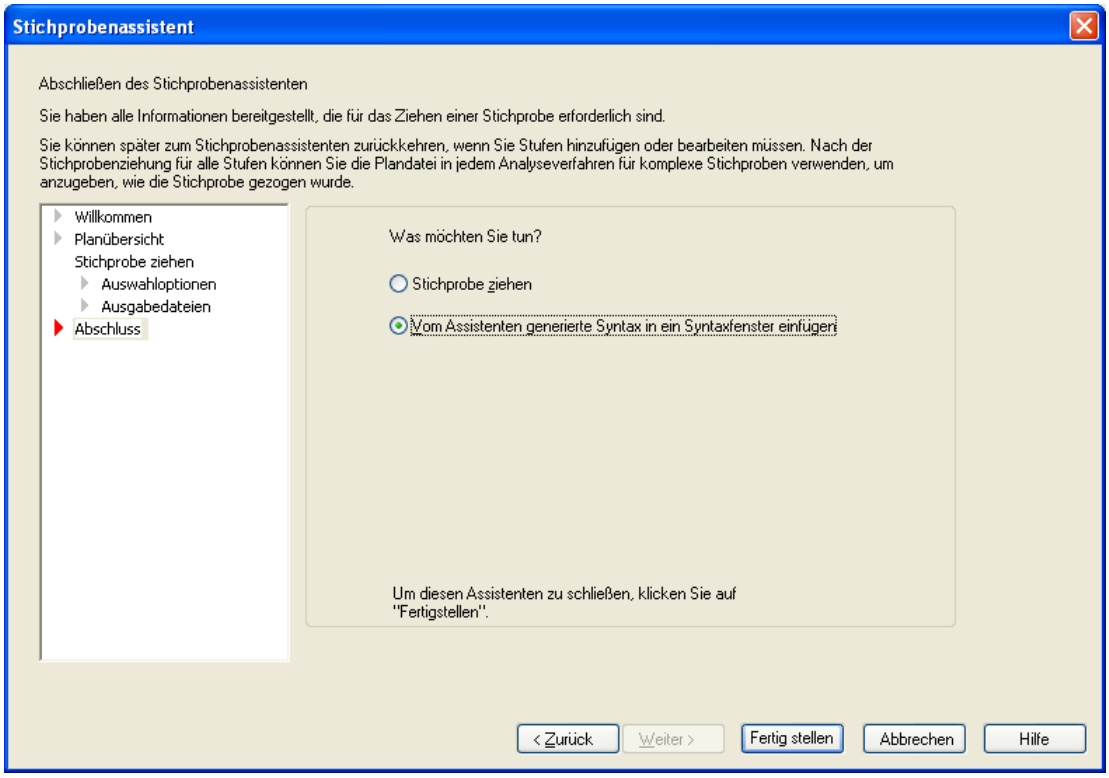
- Wählen Sie 1, 2 als Stufen, für die bereits Stichproben gezogen wurden.
- Klicken Sie auf Weiter.

Abbildung 13-29
Stichprobenassistent – Schritt "Stichprobe ziehen: Auswahloptionen"


Stichprobenassistent
Stichprobe ziehen: Auswahloptionen
In diesem Dialogfeld können Sie auswählen, welche Stufen extrahiert werden sollen, und andere Optionen für die Stichprobenziehung festlegen, beispielsweise den für die Generierung von Zufallszahlen verwendeten Startwert.
Willkommen
Planübersicht
Stichprobe ziehen
Auswahloptionen
Ausgabedateien
Abschluss
Für welche Stufen soll die Stichprobenziehung erfolgen?
Stufen: 3
Welche Art von Startwert soll verwendet werden?
Zufällig ausgewählte Zahl
Benutzerdefinierter Wert: 4231946
Geben Sie einen Startwert ein, wenn Sie die Stichprobe später reproduzieren möchten.
In den Stichprobenumfang Fälle mit benutzerdefinierten fehlenden Schichtungs- oder Klumpenvariablen aufnehmen
Arbeitsdaten sind nach Schichtungsvariablen sortiert (vorsortierte Daten können die Verarbeitung beschleunigen)
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- Wählen Sie Benutzerdefinierter Wert als Typ für den Startwert und geben Sie 4231946 als Wert ein.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Stichprobe ziehen: Ausgabedateien" ebenfalls auf Weiter.

Abbildung 13-30
Stichprobenassistent – Schritt "Fertig stellen"


Stichprobenassistent
Abschließen des Stichprobenassistenten
Sie haben alle Informationen bereitgestellt, die für das Ziehen einer Stichprobe erforderlich sind.
Sie können später zum Stichprobenassistenten zurückkehren, wenn Sie Stufen hinzufügen oder bearbeiten müssen. Nach der Stichprobenziehung für alle Stufen können Sie die Plandatei in jedem Analyseverfahren für komplexe Stichproben verwenden, um anzugeben, wie die Stichprobe gezogen wurde.
Willkommen
Planübersicht
Stichprobe ziehen
Auswahloptionen
Ausgabedateien
Abschluss
Was möchten Sie tun?
Stichprobe ziehen
Vom Assistenten generierte Syntax in ein Syntaxfenster einfügen
Um diesen Assistenten zu schließen, klicken Sie auf "Fertigstellen".
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- Wählen Sie die Option Vom Assistenten generierte Syntax in ein Syntaxfenster einfügen.
- Klicken Sie auf Fertig stellen.

Folgende Syntax wird generiert:

```
* Stichprobenassistent.
CSSELECT
 /PLAN FILE='/demo.csplan'
 /CRITERIA STAGES = 3 SEED = 4231946
 /CLASSMISSING EXCLUDE
 /DATA RENAMEVARS
 /PRINT SELECTION.
```

In diesem Fall führt das Drucken der Stichprobenübersicht zu einer unhandlichen Tabelle, die zu Problemen im Ausgabe-Viewer führt. Um die Anzeige der Stichprobenübersicht zu deaktivieren, ersetzen Sie im Unterbefehl `PRINT` den Wert `SELECTION` durch `CPS`. Führen Sie anschließend die Syntax im Syntaxfenster aus.

Mit dieser Auswahl wird eine Stichprobe gemäß der dritten Stufe des Stichprobenplans *demo.csplan* gezogen.

Stichprobenergebnisse

Abbildung 13-31
Daten-Editor mit Stichprobenergebnissen

	city	subdivision	unit	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	InclusionProbability_3_	SampleWeightCumulative_3_	SampleWeight_Final_
14	190	946	94514	,20	5,00	,10	50,00	.	.	.
15	190	946	94515	,20	5,00	,10	50,00	.	.	.
16	190	946	94516	,20	5,00	,10	50,00	,20	244,44	244,44
17	190	946	94517	,20	5,00	,10	50,00	.	.	.
18	190	946	94518	,20	5,00	,10	50,00	.	.	.
19	190	946	94519	,20	5,00	,10	50,00	.	.	.
20	190	946	94520	,20	5,00	,10	50,00	.	.	.
21	190	946	94521	,20	5,00	,10	50,00	.	.	.
22	190	946	94522	,20	5,00	,10	50,00	.	.	.
23	190	946	94523	,20	5,00	,10	50,00	.	.	.
24	190	946	94524	,20	5,00	,10	50,00	,20	244,44	244,44
25	190	946	94525	,20	5,00	,10	50,00	.	.	.
26	190	946	94526	,20	5,00	,10	50,00	.	.	.
27	190	946	94527	,20	5,00	,10	50,00	.	.	.
28	190	946	94528	,20	5,00	,10	50,00	.	.	.
29	190	946	94529	,20	5,00	,10	50,00	,20	244,44	244,44

Datenansicht | Variablenansicht

Die Ergebnisse der Stichprobenziehung werden im Daten-Editor angezeigt. Drei neue Variablen wurden in der Arbeitsdatei gespeichert. Diese stehen für die Einschlusswahrscheinlichkeiten und die kumulierten Stichprobengewichtungen für die dritte Stufe sowie für die endgültigen Stichprobengewichtungen. Diese neuen Gewichtungen berücksichtigen die während der Stichprobenziehung für die ersten beiden Stufen berechneten Gewichtungen.

- Einheiten mit Werten für diese Variablen wurden für die Stichprobe ausgewählt.
- Einheiten mit systemdefinierten fehlenden Werten für diese Variablen wurden nicht ausgewählt.

Das Unternehmen setzt seine Ressourcen nun ein, um Umfrageinformationen für die in der Stichprobe ausgewählten Haushaltseinheiten zu gewinnen. Sobald diese Umfragen vorliegen, kann die Stichprobe mit den Analyseverfahren für komplexe Stichproben verarbeitet werden. Die Stichprobenspezifikationen entnehmen Sie dem Stichprobenplan *demo.csplan*.

Stichprobenziehung mit PPS (Probability Proportional to Size; Wahrscheinlichkeit proportional zur Größe)

Abgeordnete, die in Erwägung ziehen, einen Gesetzesentwurf einzubringen, sind daran interessiert zu ermitteln, ob dieser Gesetzesantrag öffentlich unterstützt wird und in welchem Bezug die Unterstützung für den Antrag zur demografischen Struktur der Wähler steht. Die Meinungsforscher verwenden für die Erstellung und Durchführung der entsprechenden Umfragen einen komplexen Stichprobenplan.

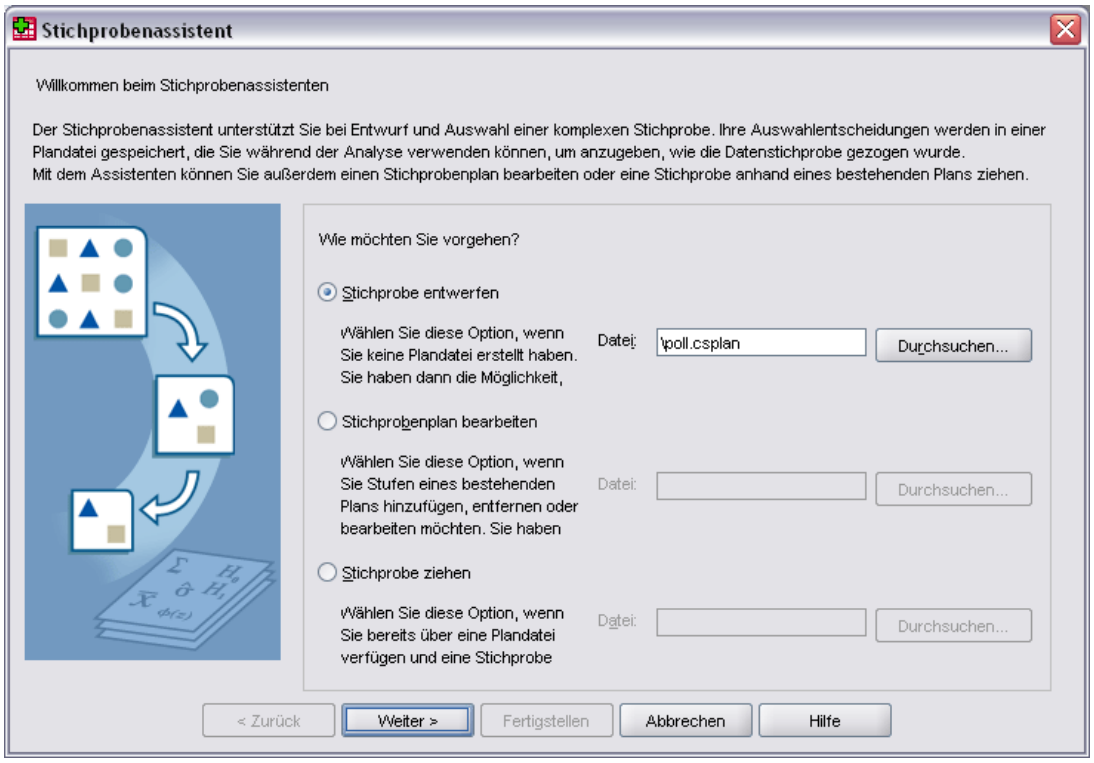
Eine Liste der registrierten Wähler finden Sie in *poll_cs.sav*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Verwenden Sie den Stichprobenassistenten für komplexe Stichproben, um eine Stichprobe zur weiteren Analyse zu ziehen.

Verwendung des Assistenten

- Um den Stichprobenassistenten für komplexe Stichproben durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

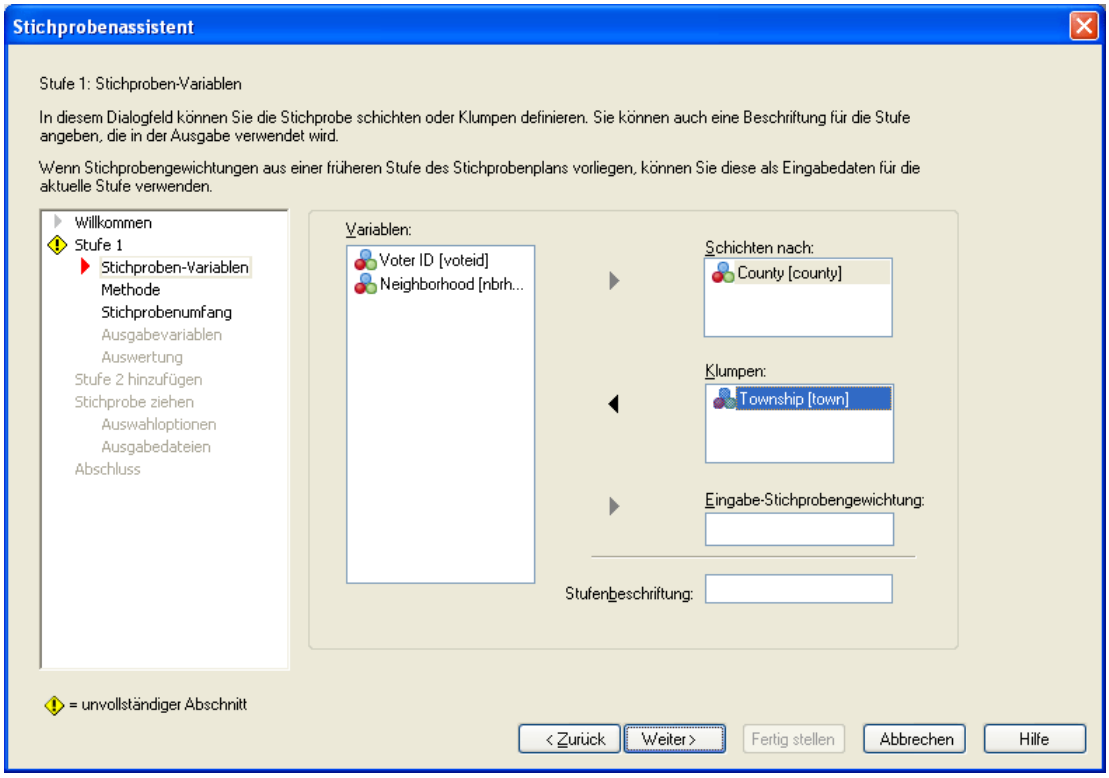
  Analysieren
  Komplexe Stichproben
  Stichprobe auswählen...

Abbildung 13-32
Stichprobenassistent – Schritt "Willkommen"


Stichprobenassistent
Willkommen beim Stichprobenassistenten
Der Stichprobenassistent unterstützt Sie bei Entwurf und Auswahl einer komplexen Stichprobe. Ihre Auswahlentscheidungen werden in einer Plandatei gespeichert, die Sie während der Analyse verwenden können, um anzugeben, wie die Datenstichprobe gezogen wurde.
Mit dem Assistenten können Sie außerdem einen Stichprobenplan bearbeiten oder eine Stichprobe anhand eines bestehenden Plans ziehen.
Wie möchten Sie vorgehen?
Stichprobe entwerfen
Wählen Sie diese Option, wenn Sie keine Plandatei erstellt haben. Sie haben dann die Möglichkeit,
Datei:
\poll.csplan
Durchsuchen...
Stichprobenplan bearbeiten
Wählen Sie diese Option, wenn Sie Stufen eines bestehenden Plans hinzufügen, entfernen oder bearbeiten möchten. Sie haben
Datei:
Durchsuchen...
Stichprobe ziehen
Wählen Sie diese Option, wenn Sie bereits über eine Plandatei verfügen und eine Stichprobe
Datei:
Durchsuchen...
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

- Wählen Sie die Option Stichprobe entwerfen und geben Sie als Namen der Plandatei /poll.csplan ein.
- Klicken Sie auf Weiter.

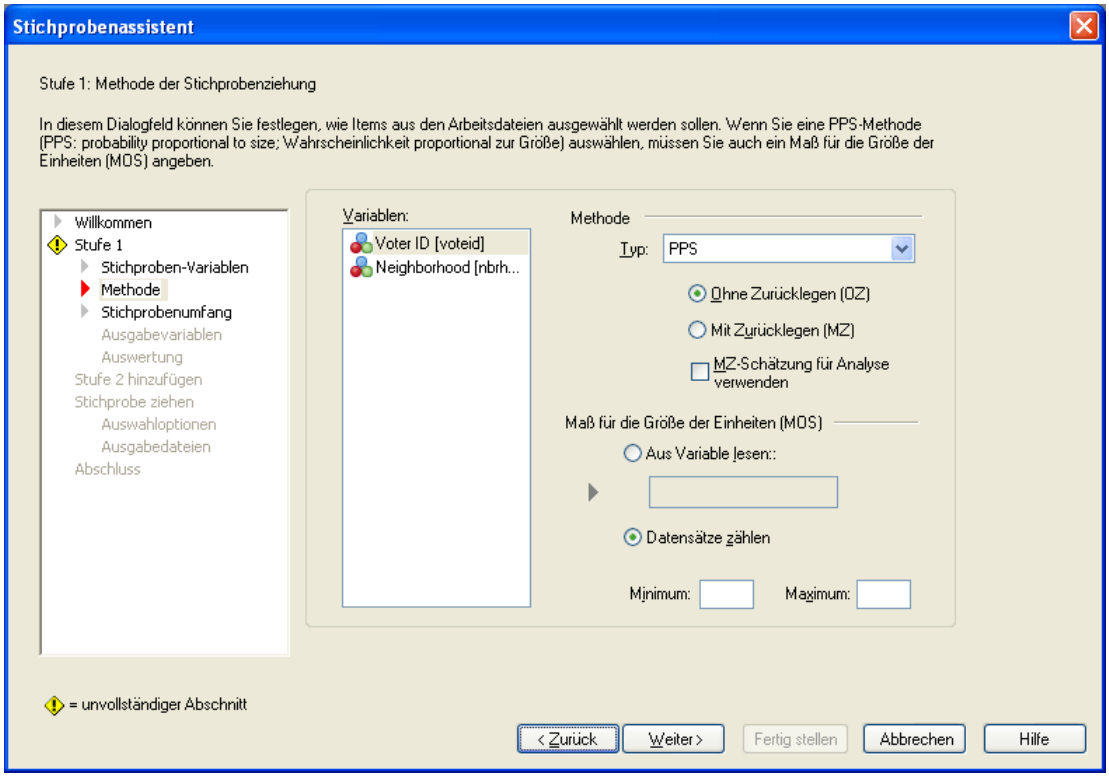
Abbildung 13-33
Stichprobenassistent – Schritt "Stichproben-Variablen" (Stufe 1)



- Wählen Sie *County* als Schichtungsvariable aus.
- Wählen Sie *Township* (Gemeinde) als Klumpenvariable aus.
- Klicken Sie auf Weiter.

Diese Planstruktur bedeutet, dass für jedes County unabhängige Stichproben gezogen werden. In dieser Phase werden "Townships" (Gemeinden) als primäre Stichprobeneinheit gezogen.

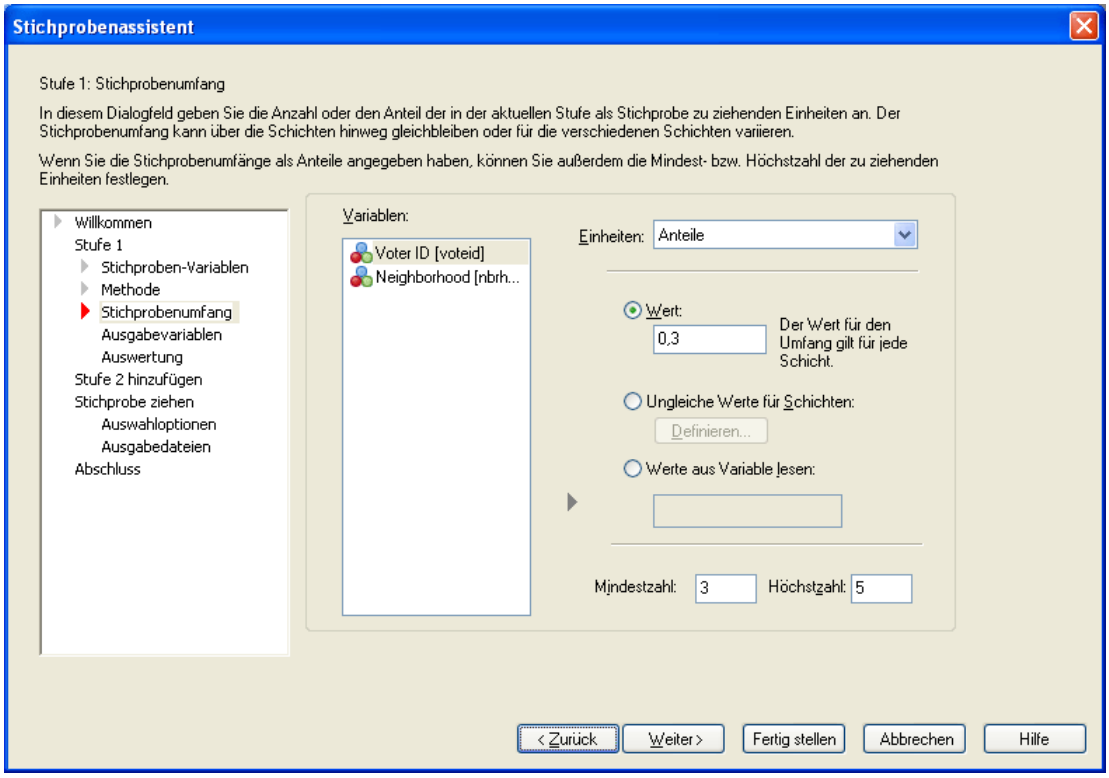
Abbildung 13-34
Stichprobenassistent – Schritt "Methode der Stichprobenziehung" (Stufe 1)



- Wählen Sie PPS als Stichprobenmethode.
- Wählen Sie Datensätze zählen als Maß für die Größe.
- Klicken Sie auf Weiter.

Innerhalb der einzelnen Counties werden Townships ohne Zurücklegen gezogen, wobei die Wahrscheinlichkeit proportional zur Anzahl der Datensätze für das jeweilige Township ist. Mit einer PPS-Methode werden gemeinsame Stichprobenwahrscheinlichkeiten für die Townships generiert. Den Speicherort für diese Werte geben Sie im Schritt "Ausgabedateien" an.

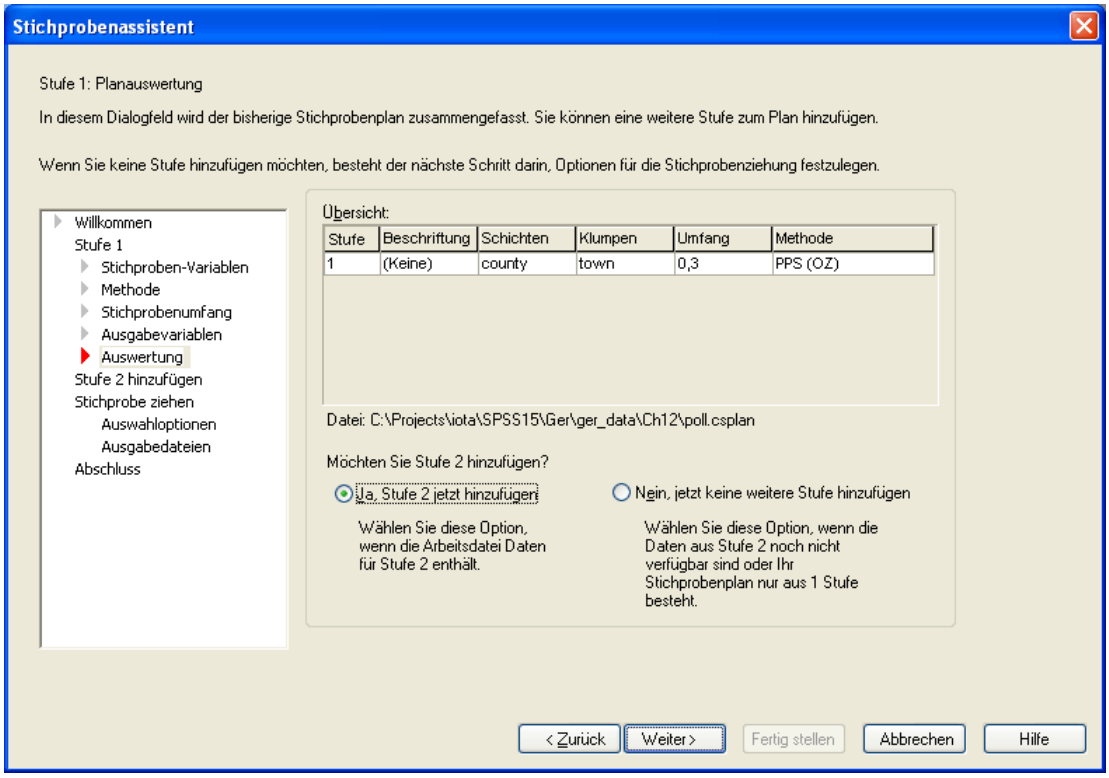
Abbildung 13-35
Stichprobenassistent, Schritt "Stichprobenumfang" (Stufe 1)



- Wählen Sie in der Dropdown-Liste für die Einheiten Anteile aus.
- Geben Sie 0,3 als Wert für den Anteil der in dieser Stufe pro County auszuwählenden Gemeinden (Townships) an.

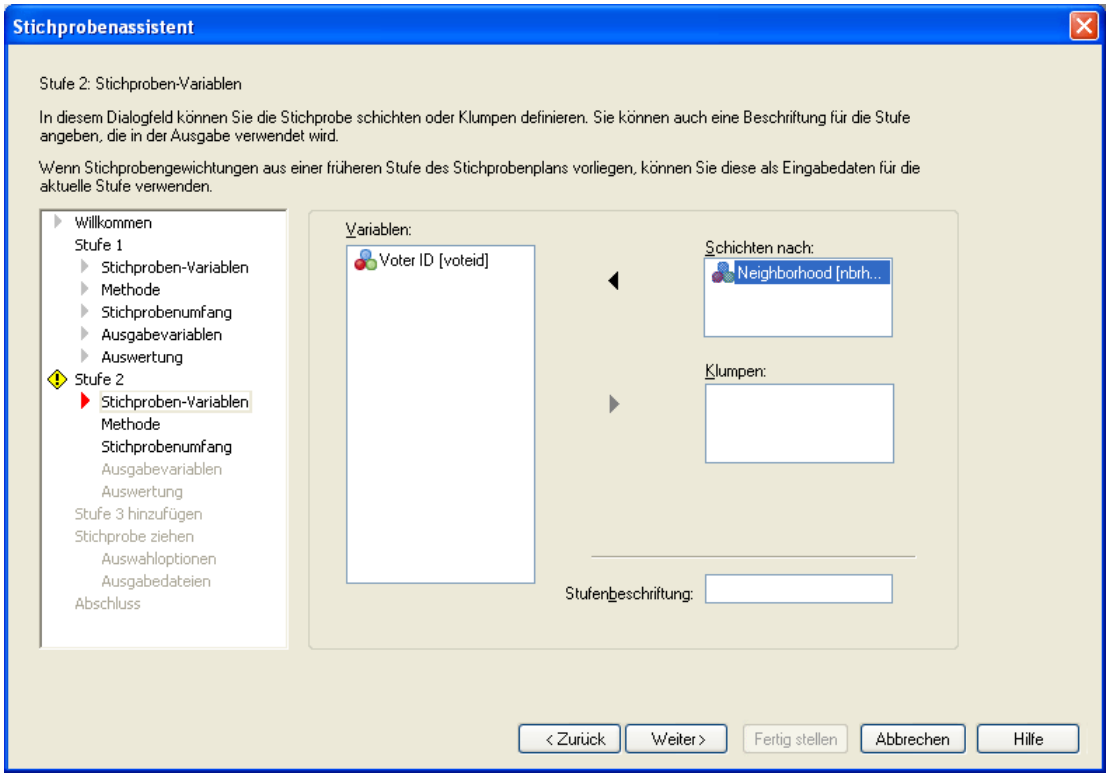
  Abgeordnete aus dem "Western County" weisen darauf hin, dass es in ihrem County weniger Gemeinden gibt als in den anderen. Um angemessen repräsentiert zu werden, würden sie gerne eine Mindeststichprobe von 3 Gemeinden für jedes County festlegen.
- Geben Sie 3 als Mindestzahl der auszuwählenden Gemeinden und 5 als Höchstzahl ein.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Ausgabevariablen" ebenfalls auf Weiter.

Abbildung 13-36
Stichprobenassistent – Schritt "Planübersicht" (Stufe 1)



- Wählen Sie Ja, Stufe 2 jetzt hinzufügen.
- Klicken Sie auf Weiter.

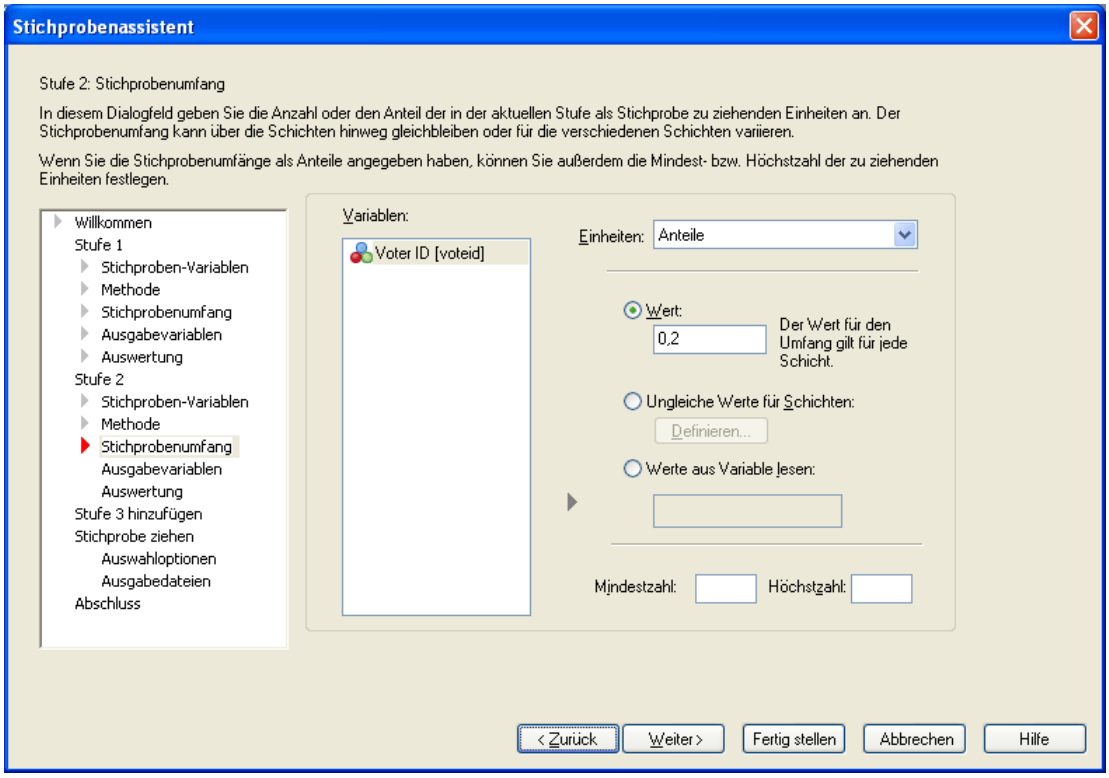
Abbildung 13-37
Stichprobenassistent – Schritt "Stichproben-Variablen" (Stufe 2)



- Wählen Sie *Neighborhood* (Wohnviertel) als Schichtungsvariable aus.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Methode der Stichprobenziehung" ebenfalls auf Weiter.

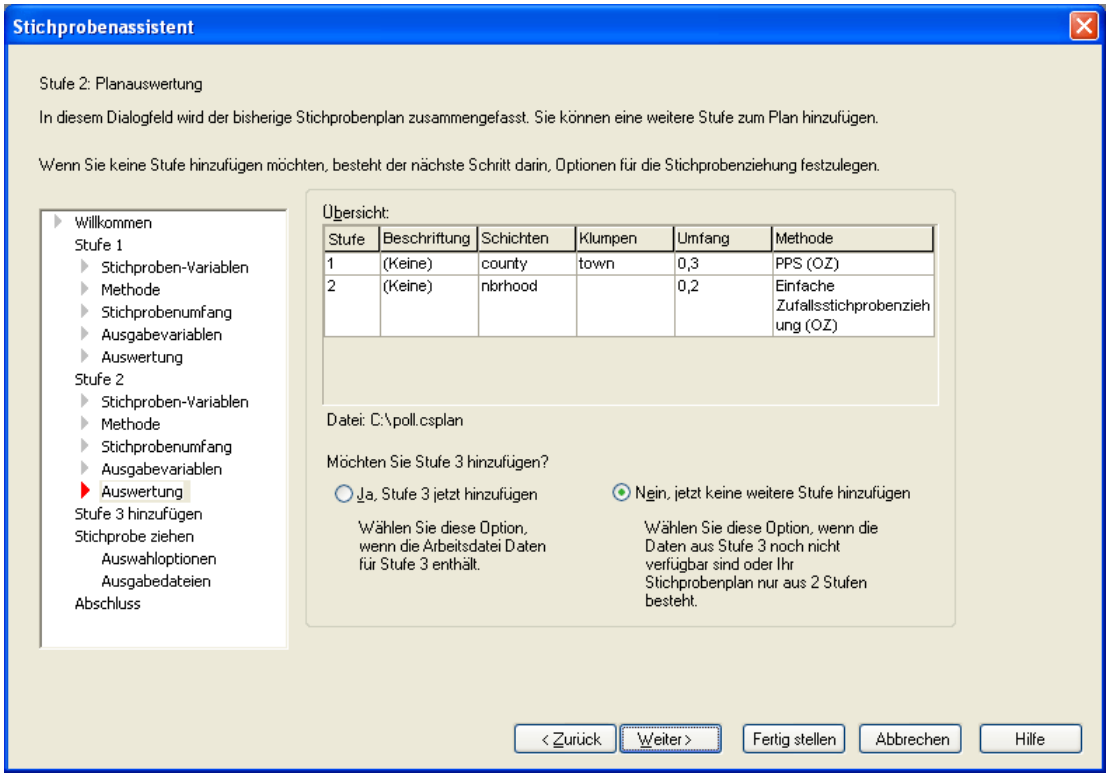
  Diese Planstruktur bedeutet, dass für jedes Wohnviertel der in Stufe 1 ausgewählten Gemeinden unabhängige Stichproben gezogen werden. In dieser Phase werden Wähler mithilfe der Methode "Einfache Zufallsstichprobenziehung" ohne Zurücklegen als primäre Stichprobeneinheit gezogen.

Abbildung 13-38
Stichprobenassistent, Schritt "Stichprobenumfang" (Stufe 2)



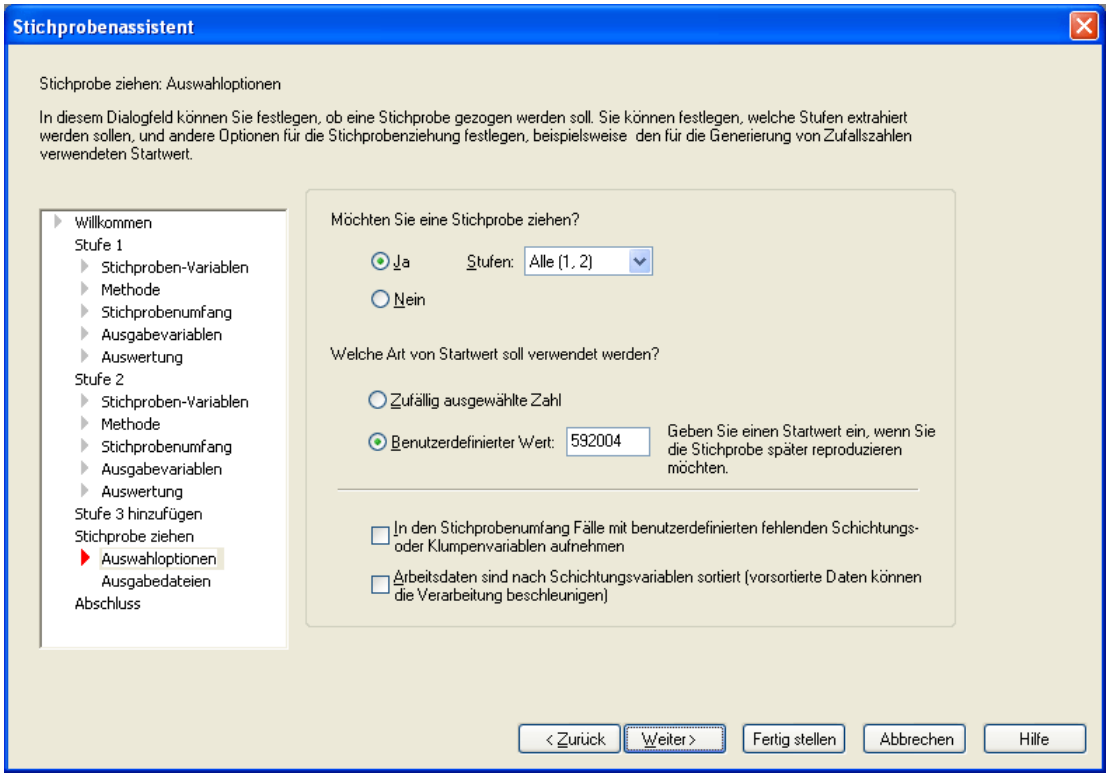
- Wählen Sie in der Dropdown-Liste für die Einheiten Anteile aus.
- Geben Sie 0,2 als Wert für den Anteil der aus jeder Schicht zu ziehenden Einheiten ein.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Ausgabevariablen" ebenfalls auf Weiter.

Abbildung 13-39
Stichprobenassistent – Schritt "Planübersicht" (Stufe 2)


Stichprobenassistent
Stufe 2: Planauswertung
In diesem Dialogfeld wird der bisherige Stichprobenplan zusammengefasst. Sie können eine weitere Stufe zum Plan hinzufügen.
Wenn Sie keine Stufe hinzufügen möchten, besteht der nächste Schritt darin, Optionen für die Stichprobenziehung festzulegen.
Willkommen
Stufe 1
Stichproben-Variablen
Methode
Stichprobenumfang
Ausgabevariablen
Auswertung
Stufe 2
Stichproben-Variablen
Methode
Stichprobenumfang
Ausgabevariablen
Auswertung
Stufe 3 hinzufügen
Stichprobe ziehen
Auswahloptionen
Ausgabedateien
Abschluss
Übersicht:
Stufe | Beschriftung | Schichten | Klumpen | Umfang | Methode
1 | (Keine) | county | town | 0,3 | PPS (OZ)
2 | (Keine) | nbrhood | | 0,2 | Einfache Zufallsstichprobenziehung (OZ)
Datei: C:\poll.csplan
Möchten Sie Stufe 3 hinzufügen?
Ja, Stufe 3 jetzt hinzufügen
Wählen Sie diese Option, wenn die Arbeitsdatei Daten für Stufe 3 enthält.
Nein, jetzt keine weitere Stufe hinzufügen
Wählen Sie diese Option, wenn die Daten aus Stufe 3 noch nicht verfügbar sind oder Ihr Stichprobenplan nur aus 2 Stufen besteht.
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- Überprüfen Sie den Stichprobenplan und klicken Sie dann auf Weiter.

Abbildung 13-40
Stichprobenassistent – Schritt "Stichprobe ziehen: Auswahloptionen"

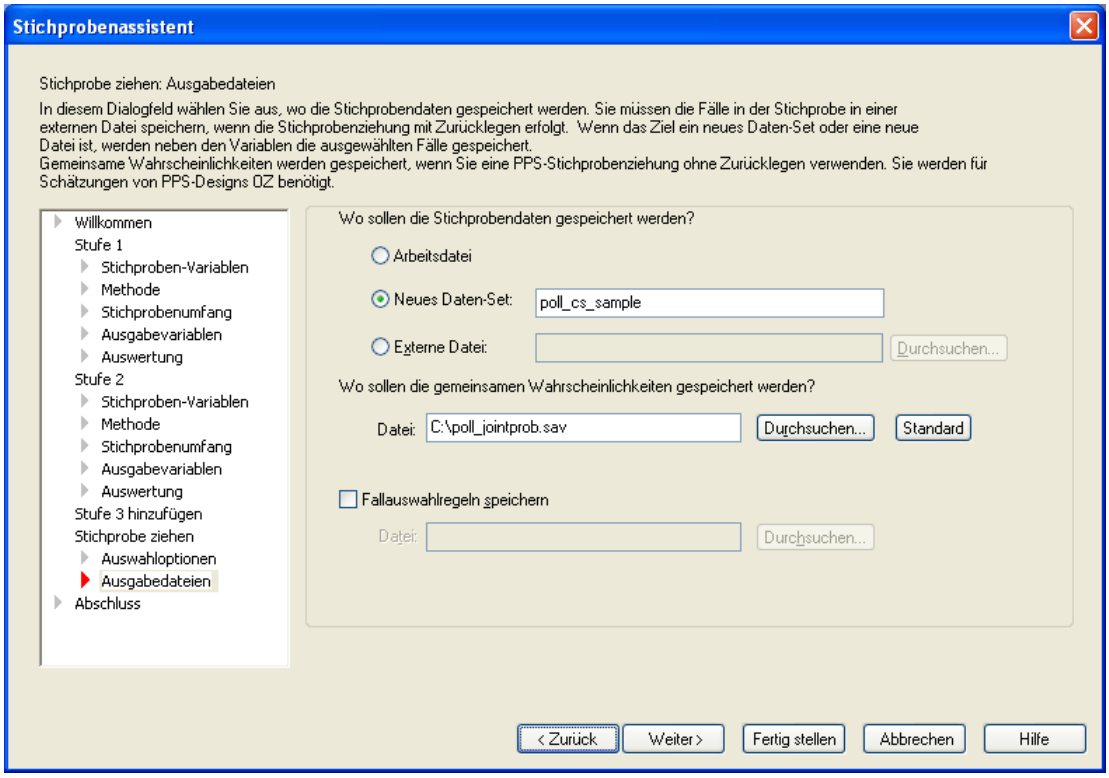


- Wählen Sie Benutzerdefinierter Wert als Typ für den Startwert und geben Sie 592004 als Wert ein.

  Durch die Verwendung eines benutzerdefinierten Werts können Sie die Ergebnisse in diesem Beispiel genau reproduzieren.

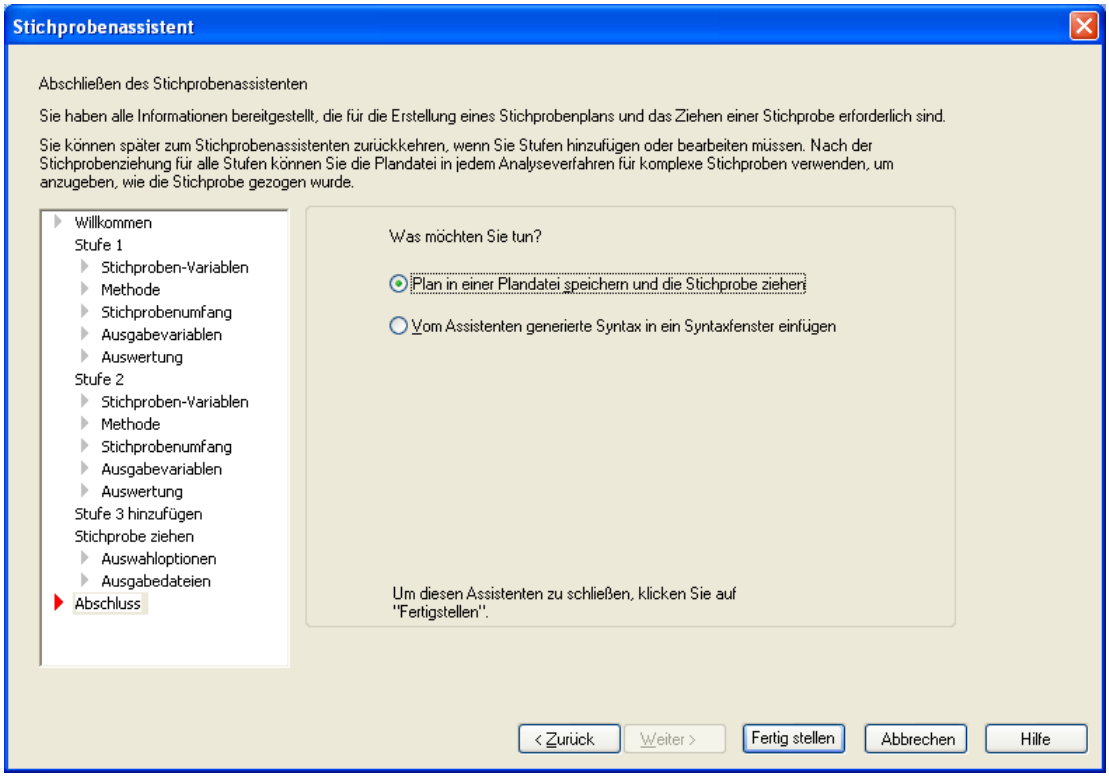
- Klicken Sie auf Weiter.

Abbildung 13-41
Stichprobenassistent – Schritt "Stichprobe ziehen: Auswahloptionen"



- Wählen Sie aus, dass die Stichprobe in einem neuen Daten-Set gespeichert werden soll, und geben Sie /poll_cs_sample als Name des Daten-Sets ein.
- Geben Sie /poll_jointprob.sav als Name der Datei für die gemeinsamen Wahrscheinlichkeiten ein.
- Klicken Sie auf Weiter.

Abbildung 13-42
Stichprobenassistent – Schritt "Fertig stellen"


Stichprobenassistent
Abschließen des Stichprobenassistenten
Sie haben alle Informationen bereitgestellt, die für die Erstellung eines Stichprobenplans und das Ziehen einer Stichprobe erforderlich sind.
Sie können später zum Stichprobenassistenten zurückkehren, wenn Sie Stufen hinzufügen oder bearbeiten müssen. Nach der Stichprobenziehung für alle Stufen können Sie die Plandatei in jedem Analyseverfahren für komplexe Stichproben verwenden, um anzugeben, wie die Stichprobe gezogen wurde.
Willkommen
Stufe 1
Stichproben-Variablen
Methode
Stichprobenumfang
Ausgabevariablen
Auswertung
Stufe 2
Stichproben-Variablen
Methode
Stichprobenumfang
Ausgabevariablen
Auswertung
Stufe 3 hinzufügen
Stichprobe ziehen
Auswahloptionen
Ausgabedateien
Abschluss
Was möchten Sie tun?
Plan in einer Plandatei speichern und die Stichprobe ziehen
Vom Assistenten generierte Syntax in ein Syntaxfenster einfügen
Um diesen Assistenten zu schließen, klicken Sie auf "Fertigstellen".
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

▶ Klicken Sie auf Fertig stellen.

Durch diese Auswahl wird die Stichprobenplan-Datei *poll.csplan* erstellt und es wird gemäß diesem Plan eine Stichprobe gezogen. Die Ergebnisse der Stichprobenziehung werden im neuen Daten-Set *poll_cs_sample* gespeichert und die Datei für die gemeinsamen Wahrscheinlichkeiten wird in der externen Datendatei *poll_jointprob.sav* gespeichert.

Planübersicht

Abbildung 13-43
Planübersicht

			Stufe 1	Stufe 2
Stichproben-Variablen	Schichtung	1	County	Neighborhood
	Klumpen	1	Township	
Informationen zur Stichprobe	Auswahlverfahren		PPS-Stichprobenziehung ohne Zurücklegen	Einfache Zufallsstichprobenziehung ohne Zurücklegen
	Maß für die Größe der Einheiten		Gewonnen aus Daten	
	Anteil der Stichprobe		,3	,2
	Minimaler Stichprobenumfang		3	
	Maximaler Stichprobenumfang		5	
	Erzeugte oder veränderte Variablen	Stufenweise Einschluss-(Auswahl-)wahrscheinlichkeiten	Inclusion Probability_1_	Inclusion Probability_2_
		Stufenweise kumulierte Stichprobengewichtung	Sample Weight Cumulative_1_	Sample Weight Cumulative_2_
Informationen für die Auswertung	Annahmen für die Schätzung		Stichprobenziehung mit ungleicher Wahrscheinlichkeit ohne Zurücklegen (unter Verwendung von gemeinsamen Einschlußwahrscheinlichkeiten)	Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen
	Einschlußwahrscheinlichkeit		Gewonnen aus Variable Inclusion Probability_1_	Gewonnen aus Variable Inclusion Probability_2_

Plandatei: C:\poll.csplan
Gewichtungsvariable: SampleWeight_Final_

Die zusammenfassende Tabelle enthält eine Übersicht über den Stichprobenplan. Anhand dieser Tabelle können Sie überprüfen, ob der Plan tatsächlich Ihren Absichten entspricht.

Stichprobenübersicht

Abbildung 13-44
Stufenübersicht

	Umfang der Stichprobe		Anteil der Stichprobe	
County	Gewünscht	Tatsächlich	Gewünscht	Tatsächlich
Eastern	4	4	30,0%	30,8%
Central	4	4	30,0%	30,8%
Western	3	3	30,0%	50,0%
Northern	5	5	30,0%	33,3%
Southern	3	3	30,0%	50,0%

Plandatei: C:\poll.csplan

Diese zusammenfassende Tabelle enthält eine Übersicht über die erste Stufe der Stichprobenziehung. Anhand dieser Tabelle können Sie überprüfen, ob die Stichprobenziehung plangemäß verlief. Erinnern Sie sich daran, dass Sie eine Stichprobe von 30 % der Gemeinden pro County angefordert haben; die tatsächlich bei der Stichprobe gezogenen Anteile liegen nahe an 30 %, außer für “Western County” und “Southern County”. Dies liegt daran, dass diese Counties jeweils nur 6 Gemeinden besitzen und Sie außerdem angegeben haben, dass mindestens drei Gemeinden pro County ausgewählt werden sollten.

Abbildung 13-45
Stufenübersicht

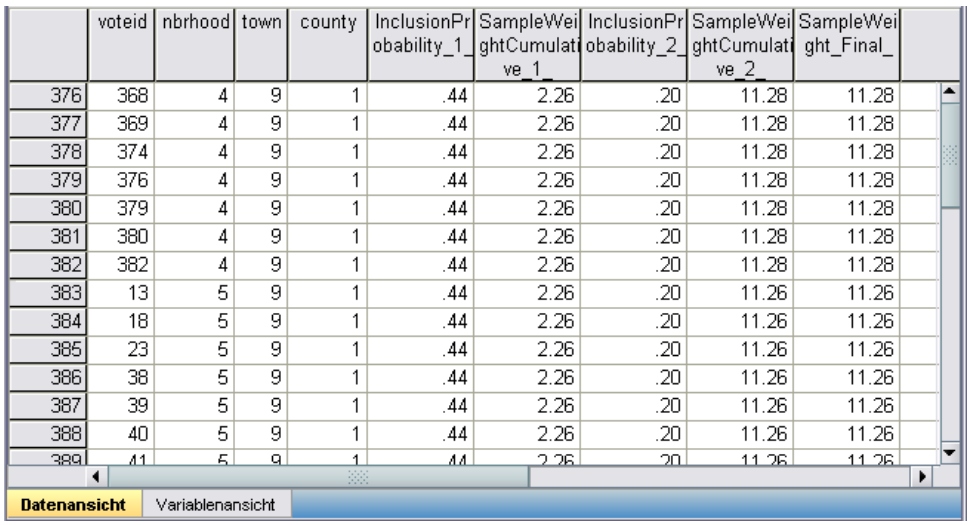
			Umfang der Stichprobe		Anteil der Stichprobe	
County	Township	Neighborhood	Gewünscht	Tatsächlich	Gewünscht	Tatsächlich
Eastern	9	1	49	49	20,0%	19,9%
		2	143	143	20,0%	20,0%
		3	113	113	20,0%	20,0%
		4	77	77	20,0%	20,0%
		5	139	139	20,0%	20,0%
		6	120	120	20,0%	20,0%
	10	1	149	149	20,0%	20,1%
		2	117	117	20,0%	20,0%
		3	116	116	20,0%	20,0%
		4	69	69	20,0%	19,9%
	11	1	65	65	20,0%	19,9%
		2	72	72	20,0%	19,9%
		3	109	109	20,0%	20,0%
		4	140	140	20,0%	20,0%
		5	42	42	20,0%	19,8%
		6	142	142	20,0%	20,0%
	12	1	145	145	20,0%	20,1%
		2	69	69	20,0%	20,1%
		3	98	98	20,0%	20,1%
		4	134	134	20,0%	20,0%
		5	114	114	20,0%	20,0%
		6	137	137	20,0%	19,9%
Central	2	1	119	119	20,0%	20,1%
		2	153	153	20,0%	19,9%
		3	101	101	20,0%	20,0%
		4	52	52	20,0%	19,8%
		5	144	144	20,0%	20,0%
	6	1	50	50	20,0%	20,1%
		2	45	45	20,0%	20,1%
		3	82	82	20,0%	20,0%
		4	107	107	20,0%	20,1%
	7	1	124	124	20,0%	20,1%
		2	77	77	20,0%	20,1%
		3	71	71	20,0%	20,0%

Plandatei: C:\poll.csplan

Diese zusammenfassende Tabelle (deren oberer Teil hier zu sehen ist) enthält eine Übersicht über die zweite Stufe der Stichprobenziehung. Außerdem können Sie damit überprüfen, ob die Stichprobenziehung plangemäß verlief. Wie angefordert, wurden ungefähr 20 % der Wähler aus jedem Wohnviertel in jeder der in der ersten Stufe ausgewählten Gemeinden als Stichprobe gezogen.

Stichprobenergebnisse

Abbildung 13-46
Daten-Editor mit Stichprobenergebnissen



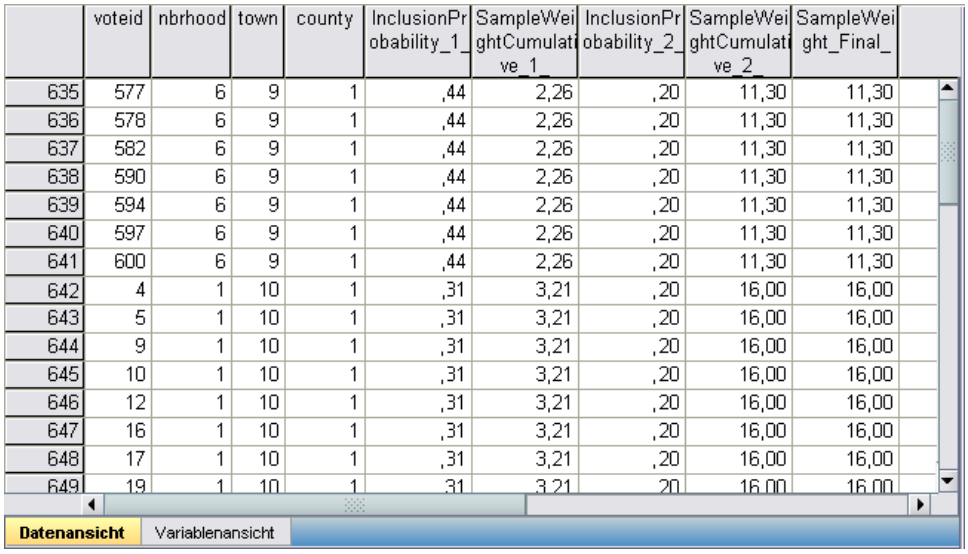
	voteid	nbrhood	town	county	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
376	368	4	9	1	.44	2.26	.20	11.28	11.28
377	369	4	9	1	.44	2.26	.20	11.28	11.28
378	374	4	9	1	.44	2.26	.20	11.28	11.28
379	376	4	9	1	.44	2.26	.20	11.28	11.28
380	379	4	9	1	.44	2.26	.20	11.28	11.28
381	380	4	9	1	.44	2.26	.20	11.28	11.28
382	382	4	9	1	.44	2.26	.20	11.28	11.28
383	13	5	9	1	.44	2.26	.20	11.26	11.26
384	18	5	9	1	.44	2.26	.20	11.26	11.26
385	23	5	9	1	.44	2.26	.20	11.26	11.26
386	38	5	9	1	.44	2.26	.20	11.26	11.26
387	39	5	9	1	.44	2.26	.20	11.26	11.26
388	40	5	9	1	.44	2.26	.20	11.26	11.26
389	41	5	9	1	.44	2.26	.20	11.26	11.26

Datenansicht | Variablenansicht

Die Ergebnisse der Stichprobenziehung werden im neu erstellten Daten-Set angezeigt. Fünf neue Variablen wurden in der Arbeitsdatei gespeichert. Diese stehen für die Einschlusswahrscheinlichkeiten und die kumulierten Stichprobengewichtungen für die einzelnen Stufen sowie für die endgültigen Stichprobengewichtungen. Wähler, die nicht für die Stichprobe ausgewählt wurden, werden aus diesem Daten-Set ausgeschlossen.

Die endgültigen Stichprobengewichte sind für Wähler im selben Wohnviertel identisch, da sie innerhalb der Wohnviertel mit einer einfachen Stichprobenziehung ausgewählt wurden. Sie sind jedoch für die Wohnviertel innerhalb derselben Gemeinde unterschiedlich, da die in der Stichprobe gezogenen Anteile nicht in allen Gemeinden genau 20 % entsprechen.

Abbildung 13-47
Daten-Editor mit Stichprobenergebnissen

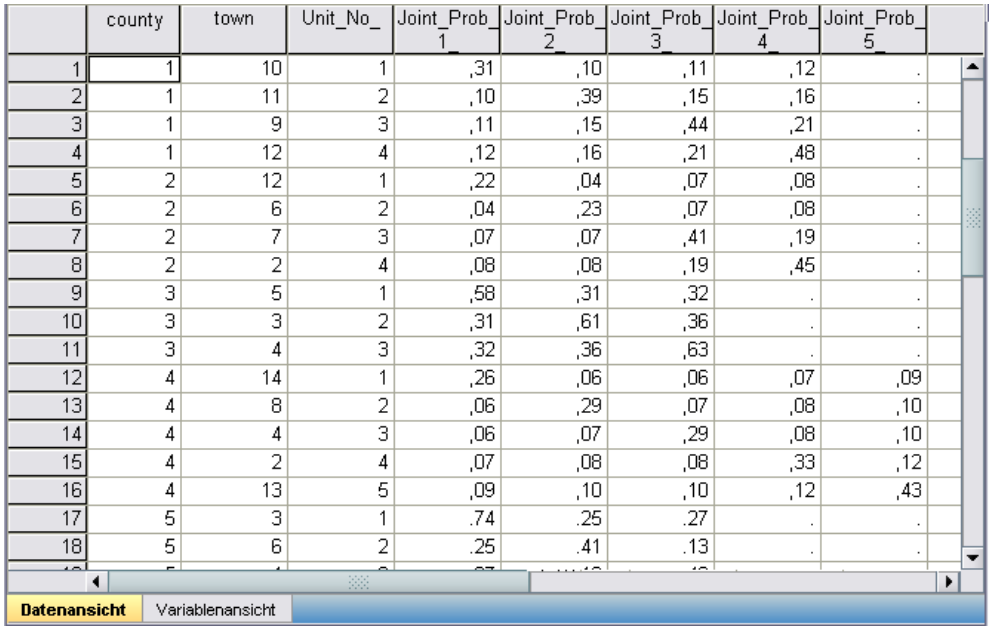


	voteid	nbrhood	town	county	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
635	577	6	9	1	,44	2,26	,20	11,30	11,30
636	578	6	9	1	,44	2,26	,20	11,30	11,30
637	582	6	9	1	,44	2,26	,20	11,30	11,30
638	590	6	9	1	,44	2,26	,20	11,30	11,30
639	594	6	9	1	,44	2,26	,20	11,30	11,30
640	597	6	9	1	,44	2,26	,20	11,30	11,30
641	600	6	9	1	,44	2,26	,20	11,30	11,30
642	4	1	10	1	,31	3,21	,20	16,00	16,00
643	5	1	10	1	,31	3,21	,20	16,00	16,00
644	9	1	10	1	,31	3,21	,20	16,00	16,00
645	10	1	10	1	,31	3,21	,20	16,00	16,00
646	12	1	10	1	,31	3,21	,20	16,00	16,00
647	16	1	10	1	,31	3,21	,20	16,00	16,00
648	17	1	10	1	,31	3,21	,20	16,00	16,00
649	19	1	10	1	,31	3,21	,20	16,00	16,00

Datenansicht | Variablenansicht

Anders als bei den Wählern in der zweiten Phase sind die Strichprobengewichte der ersten Phase nicht für Gemeinden innerhalb desselben County identisch, da sie nach der Methode “Wahrscheinlichkeit proportional zur Größe” ausgewählt wurden.

Abbildung 13-48
Datei für gemeinsame Wahrscheinlichkeiten



	county	town	Unit_No_	Joint_Prob_1_	Joint_Prob_2_	Joint_Prob_3_	Joint_Prob_4_	Joint_Prob_5_
1	1	10	1	,31	,10	,11	,12	.
2	1	11	2	,10	,39	,15	,16	.
3	1	9	3	,11	,15	,44	,21	.
4	1	12	4	,12	,16	,21	,48	.
5	2	12	1	,22	,04	,07	,08	.
6	2	6	2	,04	,23	,07	,08	.
7	2	7	3	,07	,07	,41	,19	.
8	2	2	4	,08	,08	,19	,45	.
9	3	5	1	,58	,31	,32	.	.
10	3	3	2	,31	,61	,36	.	.
11	3	4	3	,32	,36	,63	.	.
12	4	14	1	,26	,06	,06	,07	,09
13	4	8	2	,06	,29	,07	,08	,10
14	4	4	3	,06	,07	,29	,08	,10
15	4	2	4	,07	,08	,08	,33	,12
16	4	13	5	,09	,10	,10	,12	,43
17	5	3	1	.74	.25	.27	.	.
18	5	6	2	.25	.41	.13	.	.

Datenansicht | Variablenansicht

Die Datei *poll_jointprob.sav* enthält die gemeinsamen Wahrscheinlichkeiten der ersten Stufe für ausgewählte Townships in Counties. *County* Schichtungsvariable der ersten Stufe und *Township* ist eine Klumpenvariable. Durch Kombinationen aus diesen Variablen werden alle

primären Stichprobeneinheiten (Primary Sampling Units, PSUs) der ersten Stufe eindeutig identifiziert. *Unit_No_* bezeichnet PSUs in den einzelnen Schichten und wird zum Abgleich mit *Joint_Prob_1_*, *Joint_Prob_2_*, *Joint_Prob_3_*, *Joint_Prob_4_* und *Joint_Prob_5_* verwendet. Die ersten beiden Schichten weisen jeweils 4 PSUs auf. Daher haben die Matrizen für die gemeinsame Einschlusswahrscheinlichkeit die Größe 4×4 für diese Schichten und die Spalte *Joint_Prob_5_* wird für die betreffenden Zeilen leer gelassen. Die Schichten 3 und 5 weisen Matrizen der Größe 3×3 für die gemeinsame Einschlusswahrscheinlichkeit auf und Schicht 4 weist eine Matrix der Größe 5×5 für die gemeinsame Einschlusswahrscheinlichkeit auf.

Die Notwendigkeit einer Datei für gemeinsame Wahrscheinlichkeiten wird bei der Durchsicht der Werte für die Matrizen für die gemeinsame Einschlusswahrscheinlichkeit deutlich. Wenn es sich bei der Stichprobenmethode nicht um eine PPS-Methode ohne Zurücklegen handelt, ist die Auswahl einer PSU unabhängig von der Auswahl einer weiteren PSU und die gemeinsame Einschlusswahrscheinlichkeit ist einfach das Produkt der beiden Einschlusswahrscheinlichkeiten. Im Gegensatz dazu beträgt die gemeinsame Einschlusswahrscheinlichkeit für Township 9 und 10 von County 1 ungefähr 0,11 (siehe den ersten Fall von *Joint_Prob_3_* bzw. den dritten Fall von *Joint_Prob_1_*) bzw. weniger als das Produkt der einzelnen Einschlusswahrscheinlichkeiten (das Produkt des ersten Falls von *Joint_Prob_1_* und des dritten Falls von *Joint_Prob_3_* beträgt 0,31×0,44=0,1364).

Die Meinungsforscher führen nur Umfragen innerhalb der ausgewählten Stichprobe durch. Sobald die Ergebnisse vorliegen, kann die Stichprobe mit den Analyseverfahren für komplexe Stichproben verarbeitet werden. Für die Stichprobenspezifikationen wird der Stichprobenplan *poll.csplan* verwendet und für die erforderlichen gemeinsamen Einschlusswahrscheinlichkeiten die Datei *poll_jointprob.sav*.

Verwandte Prozeduren

Die Prozedur “Stichprobenassistent für komplexe Stichproben” ist ein nützliches Werkzeug zum Erstellen einer Datei für den Stichprobenplan und zum Ziehen von Stichproben.

- Um eine Stichprobe für die Analyse vorzubereiten, wenn Sie nicht auf die Datei mit dem Stichprobenplan zugreifen können, verwenden Sie den Analysevorbereitungsassistenten.

Kapitel
14

Analysevorbereitungsassistent für komplexe Stichproben

Der Analysevorbereitungsassistent führt Sie durch die Schritte zum Erstellen bzw. Bearbeiten eines Analyseplans zur Verwendung mit den verschiedenen Analyseverfahren für komplexe Stichproben. Der Assistent ist besonders nützlich, wenn Sie keinen Zugriff auf die Datei mit dem Stichprobenplan haben, der zum Ziehen der Stichprobe verwendet wurde.

Verwendung des Analysevorbereitungsassistenten für komplexe Stichproben zur Vorbereitung von öffentlich zugänglichen NHIS-Daten

"National Health Interview Survey (NHIS)" ist eine große, grundgesamtheitsbasierte Umfrage in unter der US-amerikanischen Zivilbevölkerung. Es werden persönliche Interviews in einer landesweit repräsentativen Stichprobe von Haushalten durchgeführt. Für die Mitglieder jedes Haushalts werden demografische Informationen und Beobachtungen zum Gesundheitsverhalten und Gesundheitsstatus eingeholt.

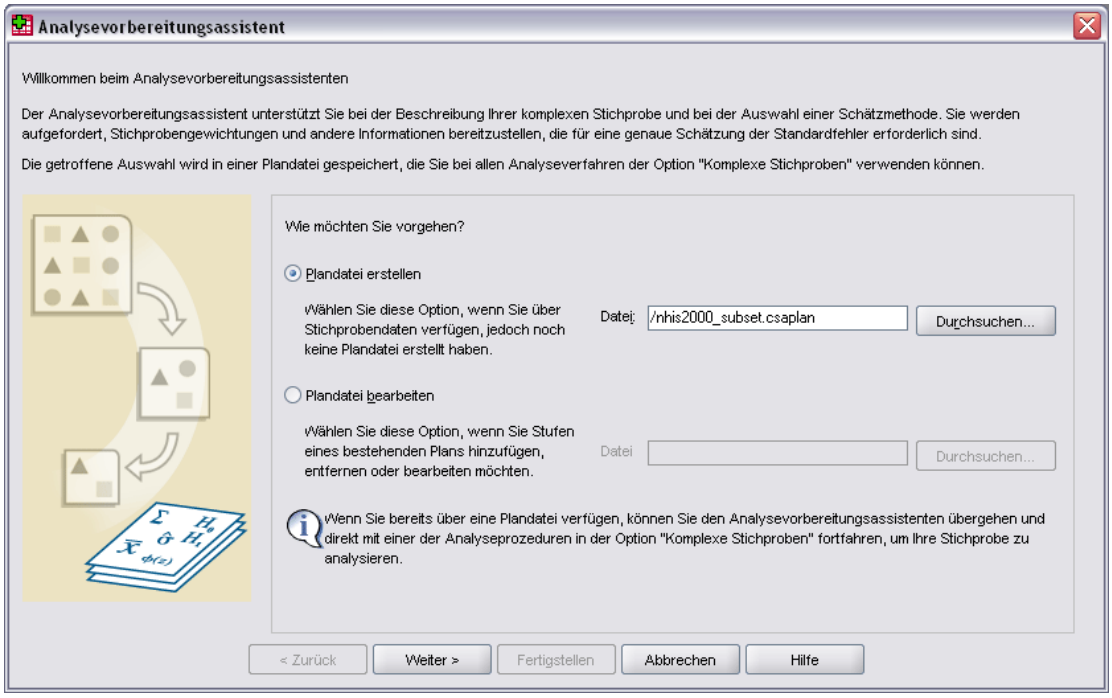
Eine Untergruppe der Umfrage aus dem Jahr 2000 finden Sie in der Datei *nhis2000_subset.sav*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Erstellen Sie mithilfe des Analysevorbereitungsassistenten für komplexe Stichproben einen Analyseplan für diese Datendatei, sodass er mit den Analyseverfahren für komplexe Stichproben verarbeitet werden kann.

Verwendung des Assistenten

- Um eine Stichprobe mithilfe des Analysevorbereitungsassistenten für komplexe Stichproben vorzubereiten, wählen Sie folgende Optionen aus den Menüs aus:

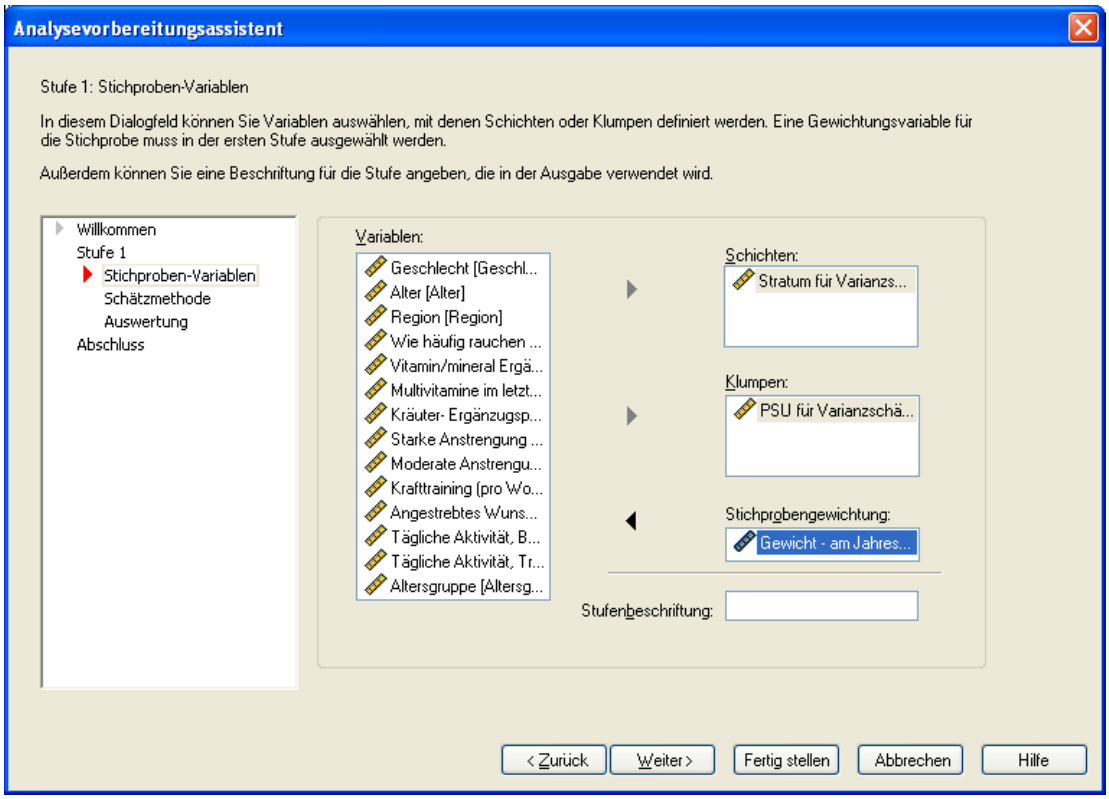
Analysieren
  Komplexe Stichproben
    Für Analyse vorbereiten...

Abbildung 14-1
Analysevorbereitungsassistent – Schritt "Willkommen"


Analysevorbereitungsassistent
Willkommen beim Analysevorbereitungsassistenten
Der Analysevorbereitungsassistent unterstützt Sie bei der Beschreibung Ihrer komplexen Stichprobe und bei der Auswahl einer Schätzmethode. Sie werden aufgefordert, Stichprobengewichtungen und andere Informationen bereitzustellen, die für eine genaue Schätzung der Standardfehler erforderlich sind.
Die getroffene Auswahl wird in einer Plandatei gespeichert, die Sie bei allen Analyseverfahren der Option "Komplexe Stichproben" verwenden können.
Wie möchten Sie vorgehen?
Plandatei erstellen
Wählen Sie diese Option, wenn Sie über Stichprobendaten verfügen, jedoch noch keine Plandatei erstellt haben.
Datei:
/nhis2000_subset.csaplan
Durchsuchen...
Plandatei bearbeiten
Wählen Sie diese Option, wenn Sie Stufen eines bestehenden Plans hinzufügen, entfernen oder bearbeiten möchten.
Datei
Durchsuchen...
Wenn Sie bereits über eine Plandatei verfügen, können Sie den Analysevorbereitungsassistenten übergehen und direkt mit einer der Analyseprozeduren in der Option "Komplexe Stichproben" fortfahren, um Ihre Stichprobe zu analysieren.
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

- Geben Sie /nhis2000_subset.csaplan als Namen für die Analyseplandatei ein.
- Klicken Sie auf Weiter.

Abbildung 14-2
Analysevorbereitungsassistent – Schritt "Stichproben-Variablen" (Stufe 1)



Die Daten werden mit einer komplexen, mehrstufigen Stichprobe gewonnen. Für die Endbenutzer wurden die ursprünglichen NHIS-Stichproben-Variablen in einen vereinfachten Satz von Stichproben- und Gewichtungsvariablen umgewandelt, deren Ergebnisse annähernd denen der ursprünglichen Stichprobenstrukturen entsprechen.

- Wählen Sie *Stratum for variance estimation* (Schicht für Varianzschätzung) als Schichtvariable.
- Wählen Sie *PSU for variance estimation* (PSU für Varianzschätzung) als Klumpenvariable.
- Wählen Sie *Weight - Final Annual* (Gewichtung - Ende jährlich) als Variable für die Stichprobengewichtung.
- Klicken Sie auf Fertig stellen.

Auswertung

Abbildung 14-3
Auswertung

			Stufe 1
Stichproben-Variablen	Schichtung	1	Stratum für Varianzschätzung
	Klumpen	1	PSU für Varianzschätzung
Informationen für die Auswertung	Annahmen für die Schätzung		Stichprobenziehung mit Zurücklegen

Plandatei: C:\nhis2000_subset.csaplan
Gewichtungsvariable: Gewicht - am Jahresende
SRS-Schätzer: Stichprobenziehung ohne Zurücklegen

In der Auswertungstabelle wird Ihr Analyseplan zusammengefasst. Der Plan besteht aus einer einzelnen Stufe mit einer (1) Schichtungsvariable und einer (1) Klumpenvariable. Die Stichprobe wird mit Zurücklegen (MZ) gezogen und der Plan wird unter *c:\nhis2000_subset.csaplan* gespeichert. Anschließend können Sie mit dieser Plandatei die Datei *nhis2000_subset.sav* mit den Analyseverfahren für komplexe Stichproben verarbeiten.

Vorbereitung für die Analyse, wenn die Datendatei keine Stichprobengewichte enthält

Eine Kreditsachbearbeiteinr verfügt über eine Sammlung von Kundendatensätzen, die anhand eines komplexen Plans zusammengestellt wurde; die Stichprobengewichte sind jedoch nicht in der Datei enthalten. Diese Informationen finden Sie in *bankloan_cs_noweights.sav*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Ausgehend von ihrem Wissen über den Stichprobenplan möchte die Sachbearbeiterin mithilfe des Analysevorbereitungsassistenten für komplexe Stichproben einen Analyseplan für diese Datendatei erstellen, der mit den Analyseverfahren für komplexe Stichproben verarbeitet werden kann.

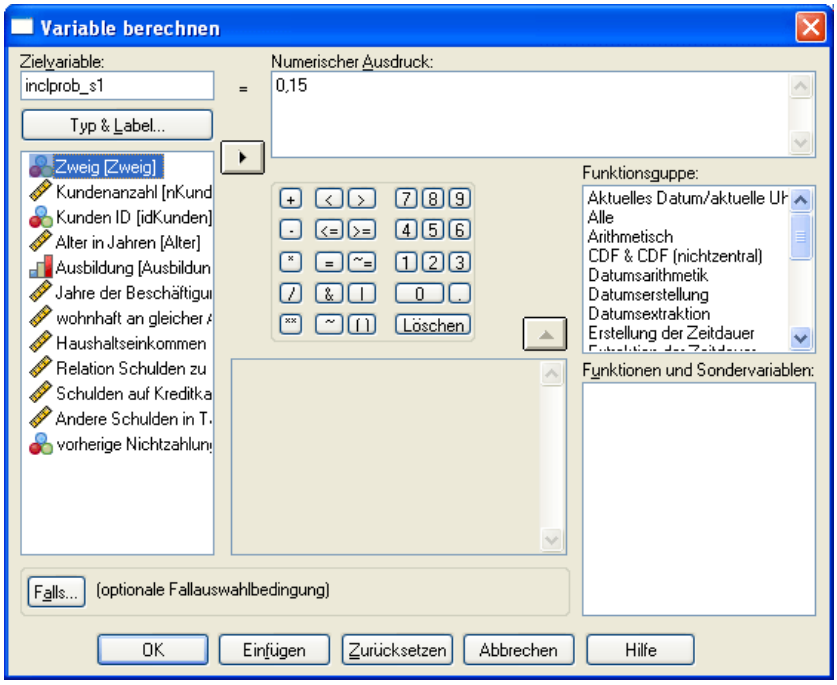
Die Kreditsachbearbeiterin weiß, dass die Datensätze in zwei Stufen ausgewählt wurden. Dabei wurden in der ersten Stufe 15 von 100 Bankfilialen mit gleicher Wahrscheinlichkeit und ohne Zurücklegen ausgewählt. In der zweiten Stufe wurden dann aus jeder dieser Banken 100 Kunden mit gleicher Wahrscheinlichkeit und ohne Zurücklegen ausgewählt und Informationen zur Anzahl der Kunden in jeder Bank sind in der Datendatei enthalten. Der erste Schritt bei der Erstellung eines Analyseplans besteht in der Berechnung der stufenweisen Einschlusswahrscheinlichkeiten und der endgültigen Stichprobengewichtungen.

Berechnung von Einschlusswahrscheinlichkeiten und Stichprobengewichten

- Um die Einschlusswahrscheinlichkeiten für die erste Stufe zu berechnen, wählen Sie folgende Befehle aus den Menüs aus:

Transformieren
  Variable berechnen...

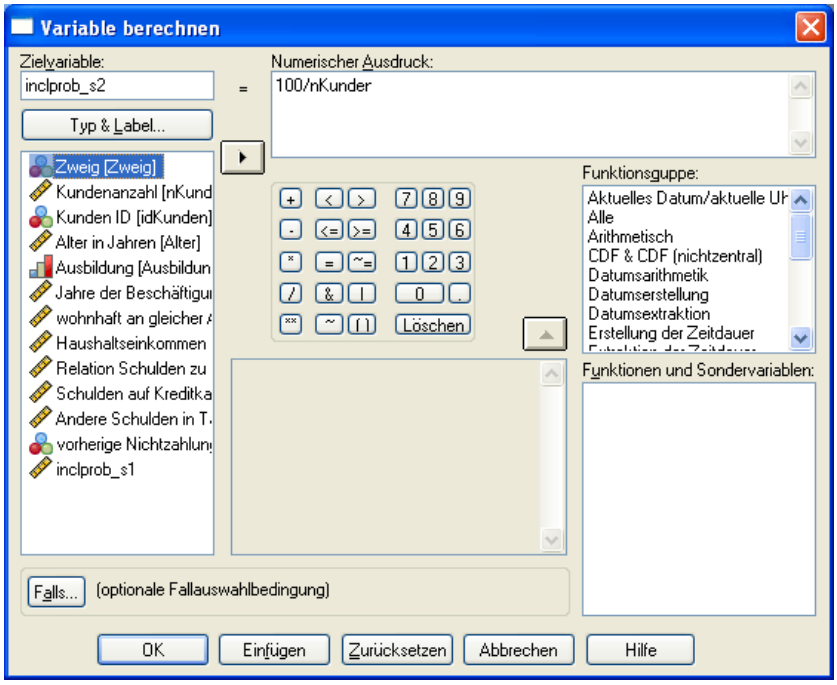
Abbildung 14-4
Dialogfeld "Variable berechnen"



In der ersten Stufe wurden 15 von 100 Bankfilialen ohne Zurücklegen ausgewählt; daher beträgt die Wahrscheinlichkeit, dass eine bestimmte Bank ausgewählt wurde, 15/100 = 0,15.

- Geben Sie inclprob_s1 als Zielvariable ein.
- Geben Sie 0.15 als numerischen Ausdruck ein.
- Klicken Sie auf OK.

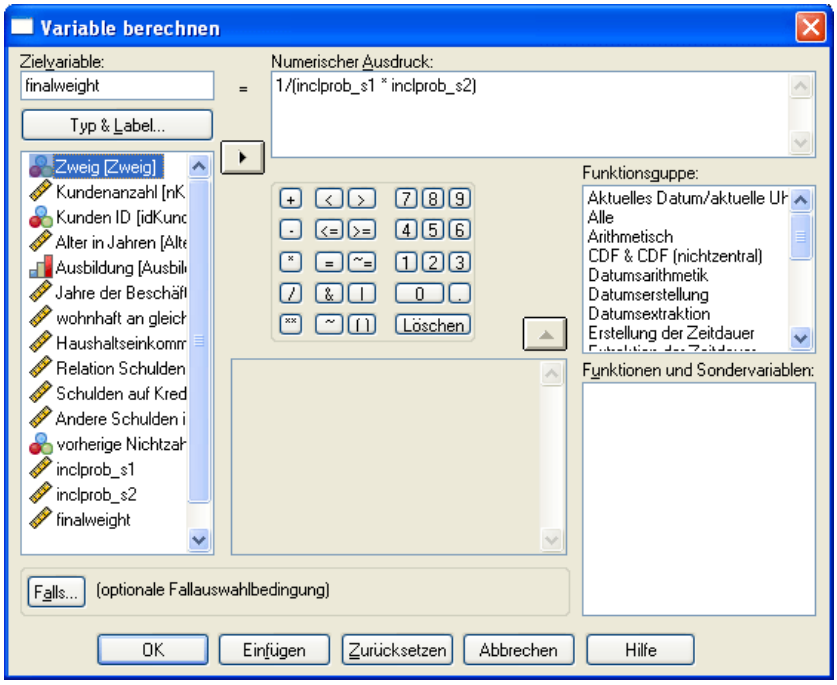
Abbildung 14-5
Dialogfeld "Variable berechnen"



In der zweiten Phase wurden aus jeder Filiale 100 Kunden ausgewählt; daher ist die Einschlusswahrscheinlichkeit in Stufe 2 für einen bestimmten Kunden in einer bestimmten Bank 100/Anzahl der Kunden dieser Bank.

- Rufen Sie das Dialogfeld "Variable berechnen" auf.
- Geben Sie inclprob_s2 als Zielvariable ein.
- Geben Sie 100/ncust als numerischen Ausdruck ein.
- Klicken Sie auf OK.

Abbildung 14-6
Dialogfeld "Variable berechnen"



Nun da die Einschlusswahrscheinlichkeiten für die einzelnen Stufen bekannt sind, können die endgültigen Stichprobengewichtungen problemlos berechnet werden.

- Rufen Sie das Dialogfeld "Variable berechnen" auf.
- Geben Sie finalweight als Zielvariable ein.
- Geben Sie 1/(inclprob_s1 * inclprob_s2)1 als numerischen Ausdruck ein.
- Klicken Sie auf OK.

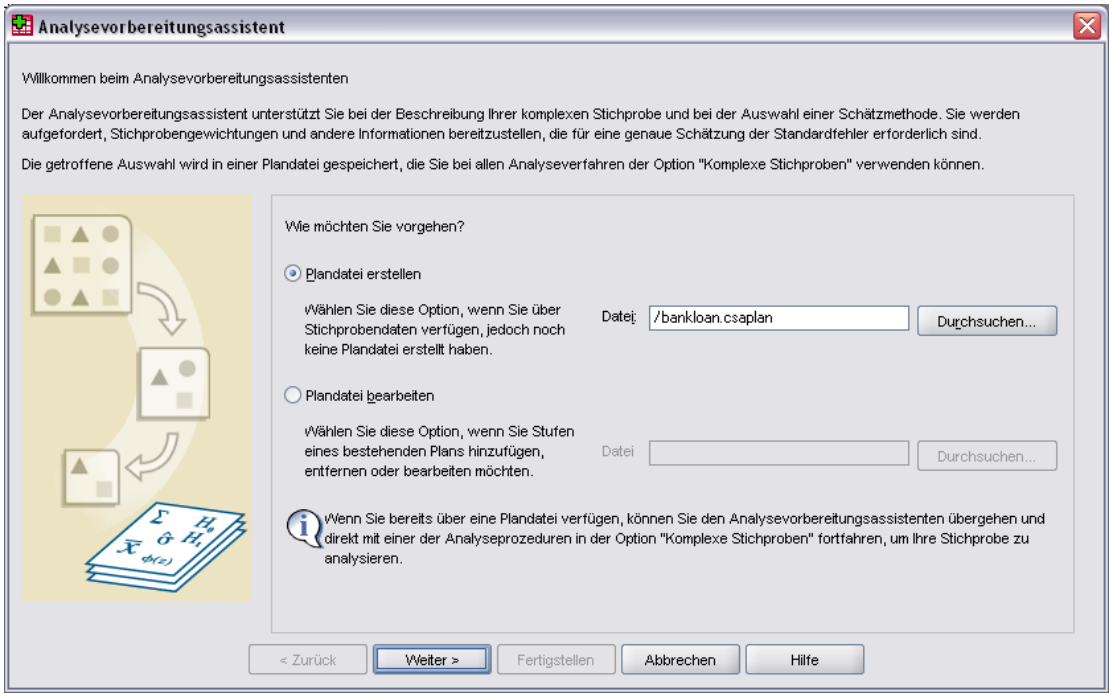
Nun können Sie den Analyseplan erstellen.

Verwendung des Assistenten

- Um eine Stichprobe mithilfe des Analysevorbereitungsassistenten für komplexe Stichproben vorzubereiten, wählen Sie folgende Optionen aus den Menüs aus:

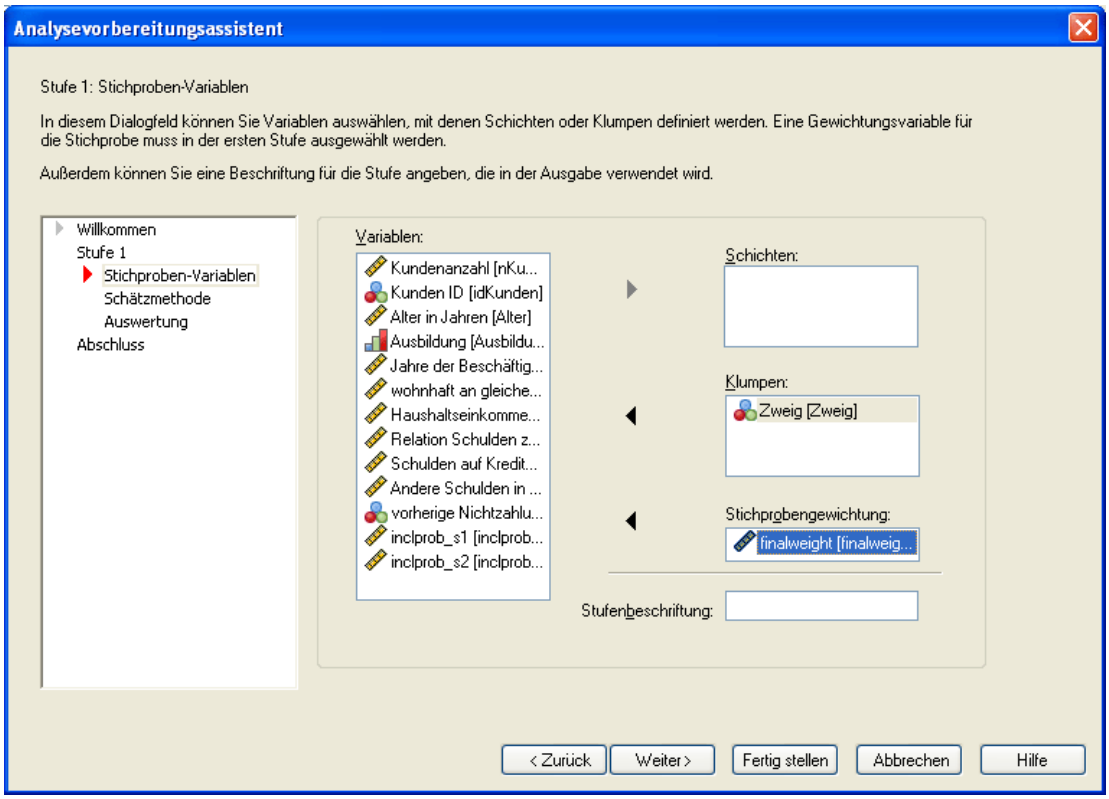
Analysieren
Komplexe Stichproben
Für Analyse vorbereiten...

Abbildung 14-7
Analysevorbereitungsassistent – Schritt "Willkommen"


Analysevorbereitungsassistent
Willkommen beim Analysevorbereitungsassistenten
Der Analysevorbereitungsassistent unterstützt Sie bei der Beschreibung Ihrer komplexen Stichprobe und bei der Auswahl einer Schätzmethode. Sie werden aufgefordert, Stichprobengewichtungen und andere Informationen bereitzustellen, die für eine genaue Schätzung der Standardfehler erforderlich sind.
Die getroffene Auswahl wird in einer Plandatei gespeichert, die Sie bei allen Analyseverfahren der Option "Komplexe Stichproben" verwenden können.
Wie möchten Sie vorgehen?
Plandatei erstellen
Wählen Sie diese Option, wenn Sie über Stichprobendaten verfügen, jedoch noch keine Plandatei erstellt haben.
Datei:
/bankloan.csaplan
Durchsuchen...
Plandatei bearbeiten
Wählen Sie diese Option, wenn Sie Stufen eines bestehenden Plans hinzufügen, entfernen oder bearbeiten möchten.
Datei
Durchsuchen...
Wenn Sie bereits über eine Plandatei verfügen, können Sie den Analysevorbereitungsassistenten übergehen und direkt mit einer der Analyseprozeduren in der Option "Komplexe Stichproben" fortfahren, um Ihre Stichprobe zu analysieren.
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

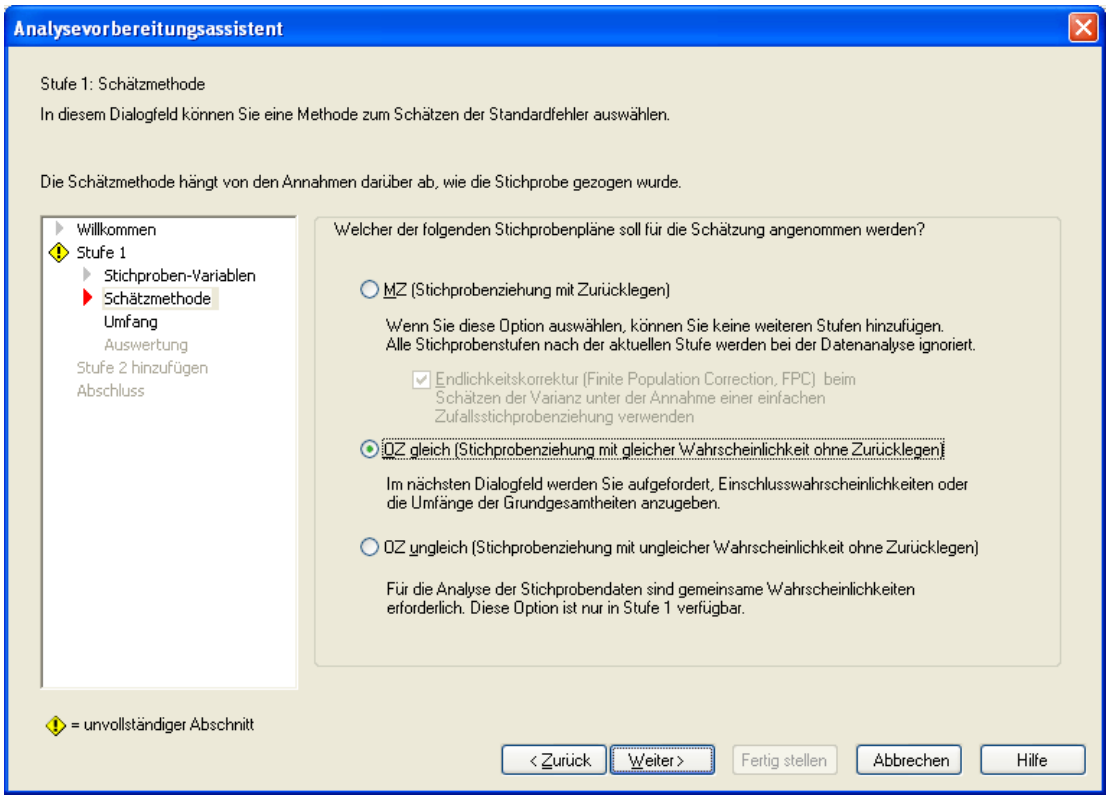
- Geben Sie /bankloan.csaplan als Namen für die Analyseplandatei ein.
- Klicken Sie auf Weiter.

Abbildung 14-8
Analysevorbereitungsassistent – Schritt "Stichproben-Variablen" (Stufe 1)



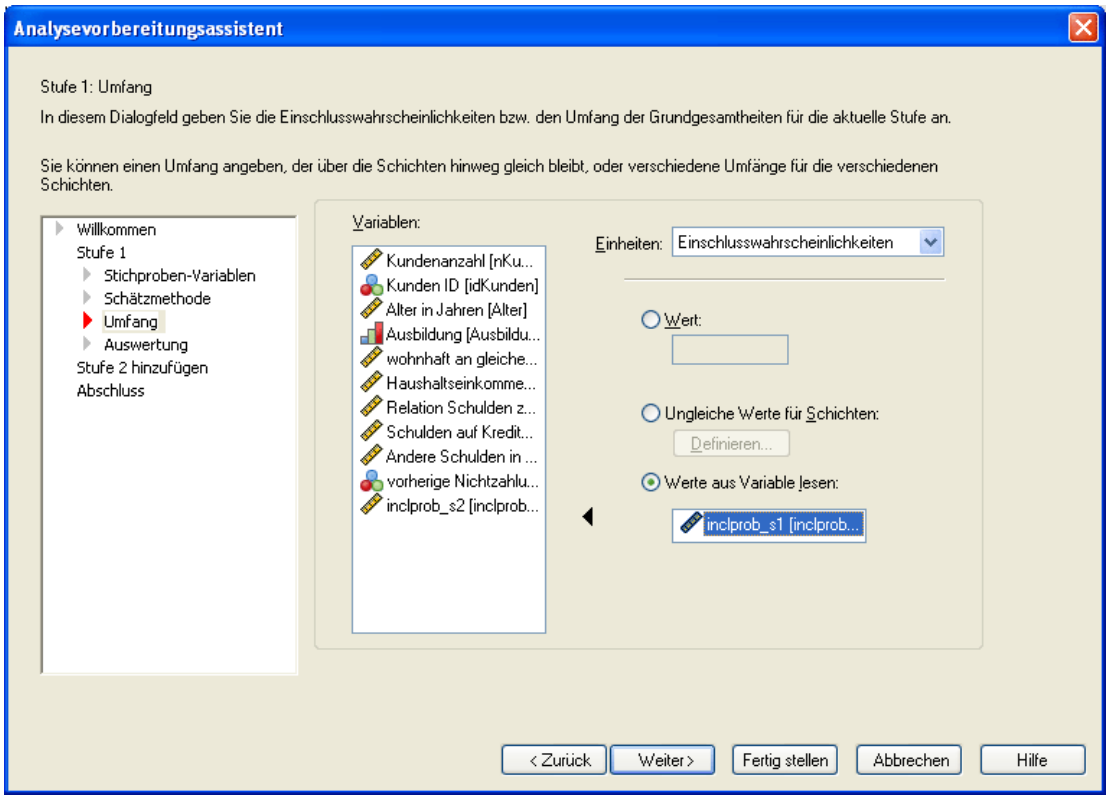
- Wählen Sie *Zweig* als Klumpenvariable aus.
- Wählen Sie *finalweight* als Variable für die Stichprobengewichtung aus.
- Klicken Sie auf Weiter.

Abbildung 14-9
Analysevorbereitungsassistent – Schritt "Schätzmethode" (Stufe 1)


Analysevorbereitungsassistent
Stufe 1: Schätzmethode
In diesem Dialogfeld können Sie eine Methode zum Schätzen der Standardfehler auswählen.
Die Schätzmethode hängt von den Annahmen darüber ab, wie die Stichprobe gezogen wurde.
Willkommen
Stufe 1
Stichproben-Variablen
Schätzmethode
Umfang
Auswertung
Stufe 2 hinzufügen
Abschluss
Welcher der folgenden Stichprobenpläne soll für die Schätzung angenommen werden?
MZ (Stichprobenziehung mit Zurücklegen)
Wenn Sie diese Option auswählen, können Sie keine weiteren Stufen hinzufügen.
Alle Stichprobenstufen nach der aktuellen Stufe werden bei der Datenanalyse ignoriert.
Endlichkeitskorrektur (Finite Population Correction, FPC) beim Schätzen der Varianz unter der Annahme einer einfachen Zufallsstichprobenziehung verwenden
OZ gleich (Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen)
Im nächsten Dialogfeld werden Sie aufgefordert, Einschlusswahrscheinlichkeiten oder die Umfänge der Grundgesamtheiten anzugeben.
OZ ungleich (Stichprobenziehung mit ungleicher Wahrscheinlichkeit ohne Zurücklegen)
Für die Analyse der Stichprobendaten sind gemeinsame Wahrscheinlichkeiten erforderlich. Diese Option ist nur in Stufe 1 verfügbar.
= unvollständiger Abschnitt
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

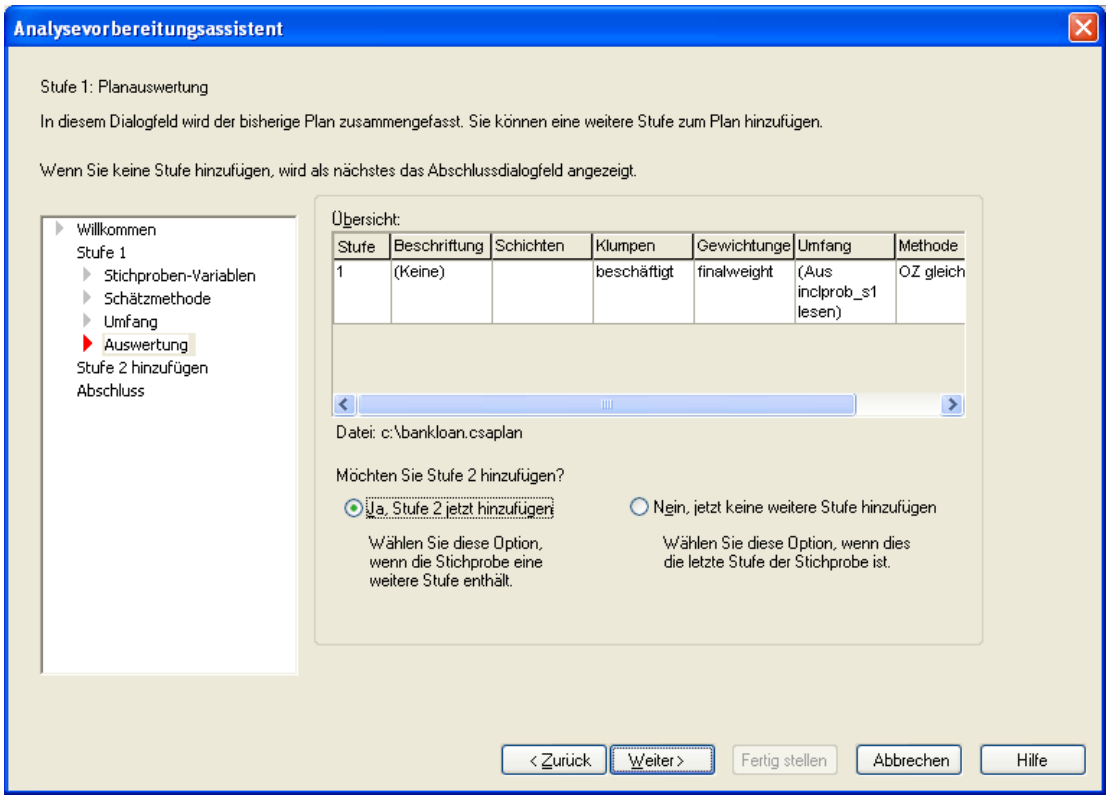
- Wählen Sie OZ gleich als Schätzmethode für die erste Stufe.
- Klicken Sie auf Weiter.

Abbildung 14-10
Analysevorbereitungsassistent – Schritt "Umfang" (Stufe 1)



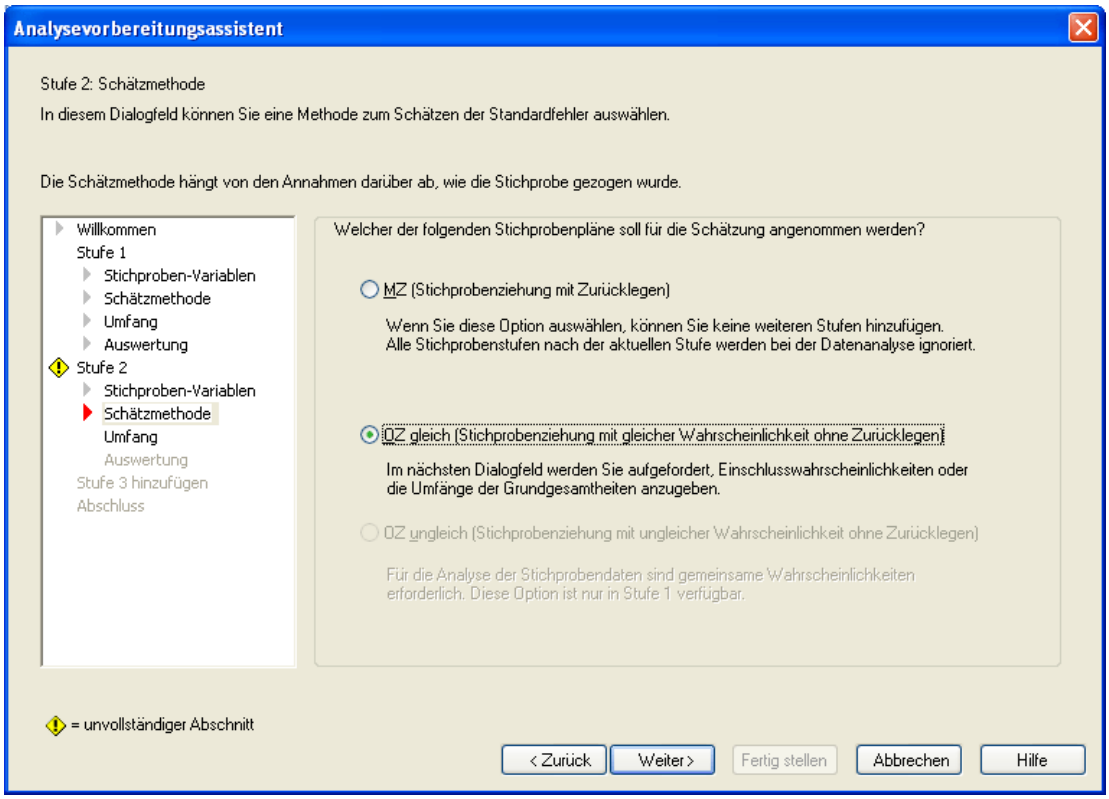
- Wählen Sie die Option Werte aus Variable lesen und wählen Sie *inclprob_s1* als die Variable aus, die die Einschlusswahrscheinlichkeiten für die erste Stufe enthält.
- Klicken Sie auf Weiter.

Abbildung 14-11
Analysevorbereitungsassistent Schritt "Planübersicht" (Stufe 1)



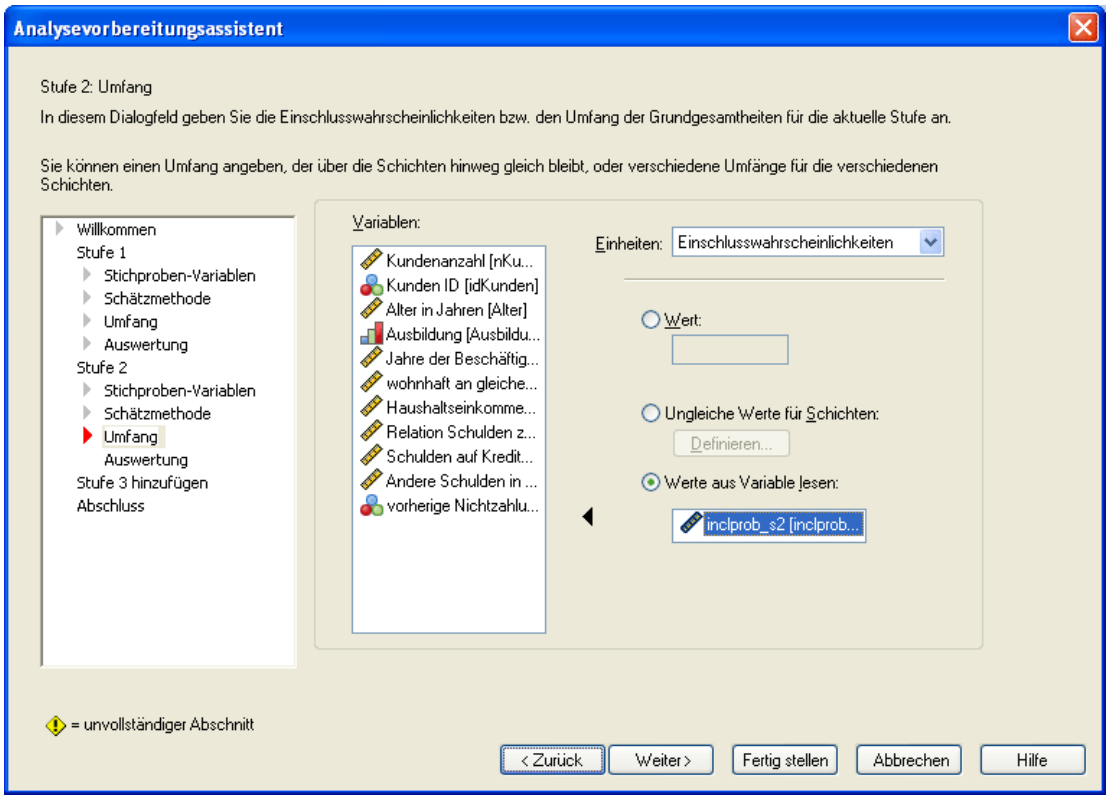
- Wählen Sie Ja, Stufe 2 jetzt hinzufügen.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Stichproben-Variablen" ebenfalls auf Weiter.

Abbildung 14-12
Analysevorbereitungsassistent – Schritt "Schätzmethode" (Stufe 2)


Analysevorbereitungsassistent
Stufe 2: Schätzmethode
In diesem Dialogfeld können Sie eine Methode zum Schätzen der Standardfehler auswählen.
Die Schätzmethode hängt von den Annahmen darüber ab, wie die Stichprobe gezogen wurde.
Willkommen
Stufe 1
Stichproben-Variablen
Schätzmethode
Umfang
Auswertung
Stufe 2
Stichproben-Variablen
Schätzmethode
Umfang
Auswertung
Stufe 3 hinzufügen
Abschluss
Welcher der folgenden Stichprobenpläne soll für die Schätzung angenommen werden?
MZ (Stichprobenziehung mit Zurücklegen)
Wenn Sie diese Option auswählen, können Sie keine weiteren Stufen hinzufügen.
Alle Stichprobenstufen nach der aktuellen Stufe werden bei der Datenanalyse ignoriert.
OZ gleich (Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen)
Im nächsten Dialogfeld werden Sie aufgefordert, Einschlusswahrscheinlichkeiten oder die Umfänge der Grundgesamtheiten anzugeben.
OZ ungleich (Stichprobenziehung mit ungleicher Wahrscheinlichkeit ohne Zurücklegen)
Für die Analyse der Stichprobendaten sind gemeinsame Wahrscheinlichkeiten erforderlich. Diese Option ist nur in Stufe 1 verfügbar.
= unvollständiger Abschnitt
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- Wählen Sie OZ gleich als Schätzmethode für die zweite Stufe.
- Klicken Sie auf Weiter.

Abbildung 14-13
Analysevorbereitungsassistent – Schritt "Umfang" (Stufe 2)



- Wählen Sie die Option Werte aus Variable lesen und wählen Sie *inclprob_s2* als die Variable aus, die die Einschlusswahrscheinlichkeiten für die zweite Stufe enthält.
- Klicken Sie auf Fertig stellen.

Auswertung

Abbildung 14-14
Zusammenfassende Tabelle

			Stufe 1	Stufe 2
Stichproben-Variablen	Klumpen	1	Zweig	
Informationen für die Auswertung	Annahmen für die Schätzung		Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen	Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen
	Einschlußwahrscheinlichkeit		Gewonnen aus Variable inclprob_s1	Gewonnen aus Variable inclprob_s2

Plandatei: C:\bankloan.csaplan
Gewichtungsvariable: finalweight
SRS-Schätzer: Stichprobenziehung ohne Zurücklegen

In der Auswertungstabelle wird Ihr Analyseplan zusammengefasst. Der Plan besteht aus zwei Stufen mit einer (1) Klumpenvariable. Für die Schätzung wird von Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen (OZ) ausgegangen und der Plan wird unter *c:\bankloan.csaplan* gespeichert. Mit dieser Plandatei können Sie nun *bankloan_noweights.sav* (mit den von Ihnen berechneten Einschlusswahrscheinlichkeiten und Stichprobengewichten) mit den Analyseverfahren für komplexe Stichproben verarbeiten.

Verwandte Prozeduren

Die Prozedur “Analysevorbereitungsassistent für komplexe Stichproben” ist ein nützliches Werkzeug für die Vorbereitung einer Stichprobe für die Analyse, wenn Sie nicht auf die Datei mit dem Stichprobenplan zugreifen können.

- Um eine Stichprobenplan-Datei zu erstellen und eine Stichprobe zu ziehen, verwenden Sie den Stichprobenassistenten.

Kapitel
15

Häufigkeiten für komplexe Stichproben

Mit der Prozedur “Häufigkeiten für komplexe Stichproben” können Sie Häufigkeitstabellen für ausgewählte Variablen erstellen und univariate Statistiken anzeigen. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Verwendung von “Häufigkeiten für komplexe Stichproben” zur Analyse der Verwendung von Nahrungsergänzungen.

Ein Forscher möchte die Verwendung von Nahrungsergänzungen bei US-Bürgern untersuchen und dafür die Ergebnisse der Umfrage “National Health Interview Survey (NHIS)” und einen zuvor erstellten Analyseplan verwenden. Für weitere Informationen siehe Verwendung des Analysevorbereitungsassistenten für komplexe Stichproben zur Vorbereitung von öffentlich zugänglichen NHIS-Daten in Kapitel 14 auf S. 148.

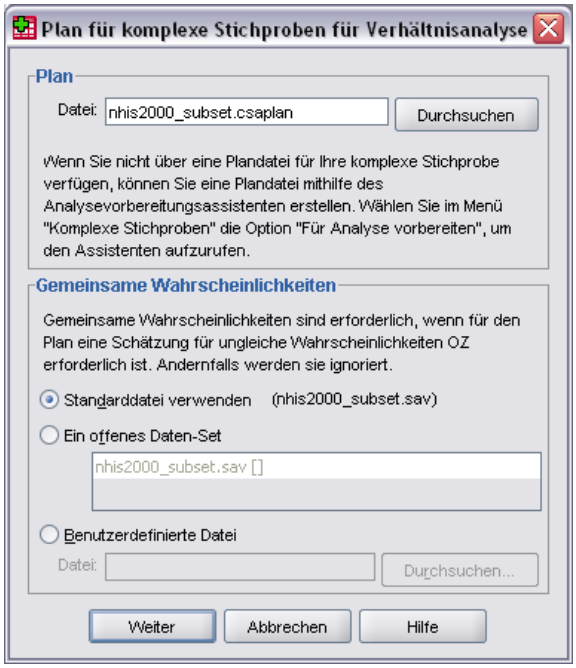
Eine Untergruppe der Umfrage aus dem Jahr 2000 finden Sie in der Datei *nhis2000_subset.sav*. Der Analyseplan ist in der Datei *nhis2000_subset.csaplan* gespeichert. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Erstellen Sie mithilfe von “Häufigkeiten für komplexe Stichproben” Statistiken für die Verwendung von Nahrungsergänzungen.

Durchführung der Analyse

▶ Um eine Analyse der Art “Häufigkeiten für komplexe Stichproben” durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

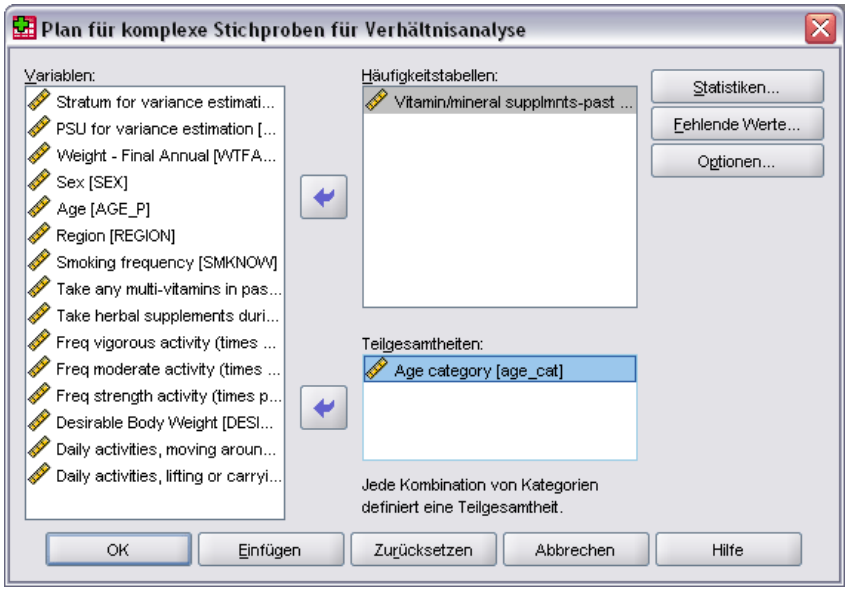
Analysieren
  Komplexe Stichproben
    Häufigkeiten...

Abbildung 15-1
Dialogfeld "Komplexe Stichproben: Plan"


Plan für komplexe Stichproben für Verhältnisanalyse
Plan
Datei:
nhis2000_subset.csaplan
Durchsuchen
Wenn Sie nicht über eine Plandatei für Ihre komplexe Stichprobe verfügen, können Sie eine Plandatei mithilfe des Analysevorbereitungsassistenten erstellen. Wählen Sie im Menü "Komplexe Stichproben" die Option "Für Analyse vorbereiten", um den Assistenten aufzurufen.
Gemeinsame Wahrscheinlichkeiten
Gemeinsame Wahrscheinlichkeiten sind erforderlich, wenn für den Plan eine Schätzung für ungleiche Wahrscheinlichkeiten OZ erforderlich ist. Andernfalls werden sie ignoriert.
Standarddatei verwenden (nhis2000_subset.sav)
Ein offenes Daten-Set
nhis2000_subset.sav []
Benutzerdefinierte Datei
Datei:
Durchsuchen...
Weiter
Abbrechen
Hilfe

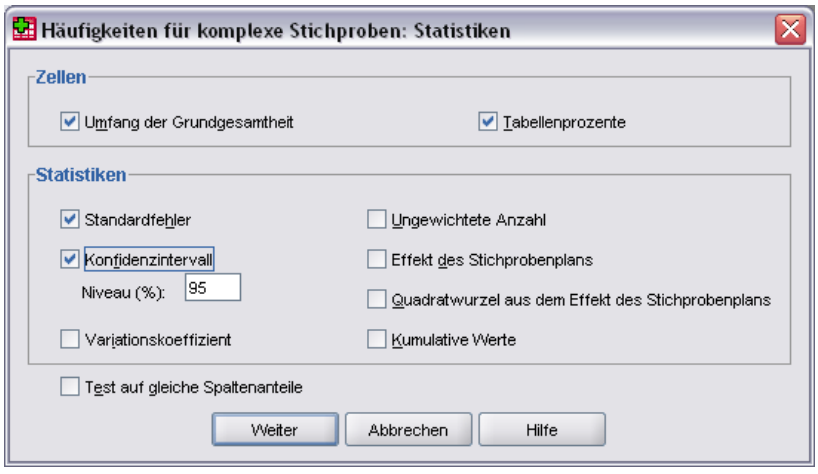
- Wechseln Sie zu der Datei *nhis2000_subset.csaplan* und wählen Sie sie aus. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271.
- Klicken Sie auf Weiter.

Abbildung 15-2
Dialogfeld "Häufigkeiten"



- Wählen Sie *Vitamin/mineral supplmnts-past 12 m* (Vitamin-/Mineralstoffergänzungen - letzte 12 Monate) als Häufigkeitsvariable aus.
- Wählen Sie *Age category* (Alterskategorie) als Teilgesamtheitsvariable aus.
- Klicken Sie auf Statistiken.

Abbildung 15-3
Dialogfeld "Häufigkeiten: Statistiken"



- Wählen Sie im Gruppenfeld "Zellen" die Option Tabellenprozente aus.
- Wählen Sie im Gruppenfeld "Statistiken" die Option Konfidenzintervall aus.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Häufigkeiten" auf OK.

Häufigkeitstabelle

Abbildung 15-4
Häufigkeitstabelle für Variable/Situation

				95%-Konfidenzintervall	
		Schätzung	Standardf ehler	Untere Grenze	Obere Grenze
Umfang der Grundgesamtheit	Ja	102767095	1185127	100435967	105098223,2
	Nein	90794234	1094402	88641560	92946907,816
	Insgesamt	193561329	1789099	190042195	197080461,9
% vom Gesamtwert	Ja	53,1%	,4%	52,4%	53,8%
	Nein	46,9%	,4%	46,2%	47,6%
	Insgesamt	100,0%	,0%	100,0%	100,0%

Jede der ausgewählten Statistiken wird für jedes der ausgewählten Zellen-Maße berechnet. Die erste Spalte enthält Schätzwerte für die Anzahl und den Prozentsatz der Personen in der Grundgesamtheit, die Vitamin-/Mineralstoffergänzungen einnehmen bzw. nicht einnehmen. Die Konfidenzintervalle überschneiden sich nicht. Daraus lässt sich ableiten, dass insgesamt die Mehrzahl der Amerikaner Vitamin-/Mineralstoffergänzungen einnimmt.

Häufigkeit nach Teilgesamtheit

Abbildung 15-5
Häufigkeitstabelle nach Teilgesamtheit

Altersgruppe			Schätzung	Standardfehler	95%-Konfidenzintervall	
					Untere Grenze	Obere Grenze
18-24	Umfang der Grundgesamtheit	Ja	10018312	350602,4	9328682	10707942,079
		Nein	15472368	499182,4	14490483	16454253,004
		Insgesamt	25490680	680732,8	24151688	26829672,224
	% vom Gesamtwert	Ja	39,3%	1,0%	37,4%	41,2%
		Nein	60,7%	1,0%	58,8%	62,6%
		Insgesamt	100,0%	,0%	100,0%	100,0%
25-44	Umfang der Grundgesamtheit	Ja	39163840	660855,7	37863946	40463734,252
		Nein	39503150	645934,2	38232606	40773693,800
		Insgesamt	78666990	961114,3	76776491	80557488,865
	% vom Gesamtwert	Ja	49,8%	,6%	48,7%	50,9%
		Nein	50,2%	,6%	49,1%	51,3%
		Insgesamt	100,0%	,0%	100,0%	100,0%
45-64	Umfang der Grundgesamtheit	Ja	34154952	598603,7	32977507	35332397,428
		Nein	24005512	497723,8	23026496	24984528,041
		Insgesamt	58160464	814680,4	56557999	59762929,346
	% vom Gesamtwert	Ja	58,7%	,6%	57,5%	60,0%
		Nein	41,3%	,6%	40,0%	42,5%
		Insgesamt	100,0%	,0%	100,0%	100,0%
65+	Umfang der Grundgesamtheit	Ja	19429991	439459,8	18565580	20294402,464
		Nein	11813204	314238,1	11195102	12431306,045
		Insgesamt	31243195	587623,4	30087348	32399042,348
	% vom Gesamtwert	Ja	62,2%	,7%	60,7%	63,6%
		Nein	37,8%	,7%	36,4%	39,3%
		Insgesamt	100,0%	,0%	100,0%	100,0%

Beim Berechnen von Statistiken nach Teilgesamtheit wird jede ausgewählte Statistik für jedes ausgewählte Zellen-Maß anhand des Werts *Age category* (Altersgruppe) berechnet. Die erste Spalte enthält Schätzwerte für die Anzahl und den Prozentsatz der Personen in jeder Kategorie, die Vitamin-/Mineralstoffergänzungen einnehmen bzw. nicht einnehmen. Die Konfidenzintervalle für die Tabellenprozentsätze überschneiden sich nicht. Daraus lässt sich ableiten, dass die Einnahme von Vitamin-/Mineralstoffergänzungen mit steigendem Alter zunimmt.

Zusammenfassung

Mit der Prozedur “Häufigkeiten für komplexe Stichproben” haben Sie Statistiken für die Verwendung von Nahrungsergänzungen bei US-Bürgern erhalten.

- Insgesamt nimmt die Mehrzahl der Amerikaner Vitamin-/Mineralstoffergänzungen ein.
- Aufgeschlüsselt nach Alter lässt sich sagen, dass mit zunehmendem Alter ein höherer Prozentsatz der Amerikaner Vitamin-/Mineralstoffergänzungen einnimmt.

Verwandte Prozeduren

Die Prozedur “Häufigkeiten für komplexe Stichproben” ist ein nützliches Werkzeug zur Gewinnung deskriptiver Statistiken von kategorialen Variablen für Beobachtungen, die mittels eines komplexen Stichprobenplans gewonnen wurden.

- Der Stichprobenassistent für komplexe Stichproben wird zur Angabe der Planspezifikationen für komplexe Stichproben und zum Ziehen von Stichproben verwendet. Die vom Stichprobenassistenten erstellte Stichprobenplan-Datei enthält einen Standard-Analyseplan und kann im Dialogfeld “Plan” angegeben werden, wenn die gezogene Stichprobe gemäß diesem Plan analysiert werden soll.
- Der Analysevorbereitungsassistent für komplexe Stichproben wird zur Angabe der Analysespezifikationen für eine bestehende komplexe Stichprobe verwendet. Die vom Stichprobenassistenten erstellte Analyseplan-Datei kann im Dialogfeld “Plan” angegeben werden, wenn Sie die Stichprobe gemäß diesem Plan analysieren.
- Die Prozedur Kreuztabellen für komplexe Stichproben bietet deskriptive Statistiken für Kreuztabellen mit kategorialen Variablen.
- Die Prozedur Deskriptive Statistiken für komplexe Stichproben bietet univariate deskriptive Statistiken für metrische Variablen.

Kapitel
16

Deskriptive Statistiken für komplexe Stichproben

Die Prozedur “Deskriptive Statistiken für komplexe Stichproben” zeigt univariate Auswertungsstatistiken für verschiedene Variablen an. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Verwendung von “Deskriptive Statistiken für komplexe Stichproben” zur Analyse von Aktivitätsniveaus.

Ein Forscher möchte das Aktivitätsniveau von US-Bürgern untersuchen und dafür die Ergebnisse der Umfrage “National Health Interview Survey (NHIS)” und einen zuvor erstellten Analyseplan verwenden. Für weitere Informationen siehe Verwendung des Analysevorbereitungsassistenten für komplexe Stichproben zur Vorbereitung von öffentlich zugänglichen NHIS-Daten in Kapitel 14 auf S. 148.

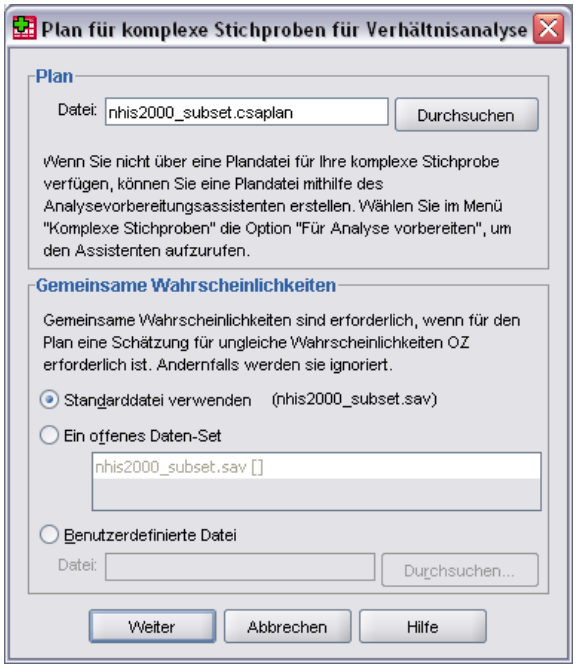
Eine Untergruppe der Umfrage aus dem Jahr 2000 finden Sie in der Datei *nhis2000_subset.sav*. Der Analyseplan ist in der Datei *nhis2000_subset.csaplan* gespeichert. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Mit “Deskriptive Statistiken für komplexe Stichproben” können Sie univariate deskriptive Statistiken für Aktivitätsniveaus erstellen.

Durchführung der Analyse

▶ Um eine Analyse der Art “Deskriptive Statistiken für komplexe Stichproben” durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

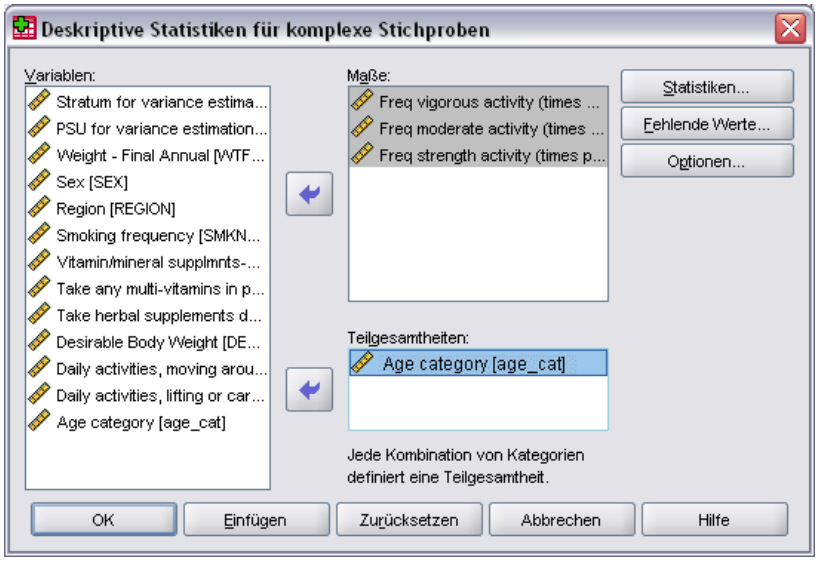
Analysieren
  Komplexe Stichproben
    Deskriptive Statistiken...

Abbildung 16-1
Dialogfeld "Komplexe Stichproben: Plan"


Plan für komplexe Stichproben für Verhältnisanalyse
Plan
Datei:
nhis2000_subset.csaplan
Durchsuchen
Wenn Sie nicht über eine Plandatei für Ihre komplexe Stichprobe verfügen, können Sie eine Plandatei mithilfe des Analysevorbereitungsassistenten erstellen. Wählen Sie im Menü "Komplexe Stichproben" die Option "Für Analyse vorbereiten", um den Assistenten aufzurufen.
Gemeinsame Wahrscheinlichkeiten
Gemeinsame Wahrscheinlichkeiten sind erforderlich, wenn für den Plan eine Schätzung für ungleiche Wahrscheinlichkeiten OZ erforderlich ist. Andernfalls werden sie ignoriert.
Standarddatei verwenden (nhis2000_subset.sav)
Ein offenes Daten-Set
nhis2000_subset.sav []
Benutzerdefinierte Datei
Datei:
Durchsuchen...
Weiter
Abbrechen
Hilfe

- Wechseln Sie zu der Datei *nhis2000_subset.csaplan* und wählen Sie sie aus. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271.
- Klicken Sie auf Weiter.

Abbildung 16-2
Dialogfeld "Deskriptive Statistik"



- Wählen Sie *Freq vigorous activity (times per wk)* (Häufigkeit der intensiven körperlichen Bewegung (pro Woche)) bis *Freq strength activity (times per wk)* (Häufigkeit des Krafttrainings (pro Woche)) als Maßvariablen aus.
- Wählen Sie *Age category* (Alterskategorie) als Teilgesamtheitsvariable aus.
- Klicken Sie auf Statistik.

Abbildung 16-3
Dialogfeld für deskriptive Statistiken

- Wählen Sie im Gruppenfeld "Statistiken" die Option Konfidenzintervall aus.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Deskriptive Statistiken für komplexe Stichproben" auf OK.

Univariate Statistiken

Abbildung 16-4
Univariate Statistiken

		Schätzung	Standard-fehler	95%-Konfidenzintervall	
				Untere Grenze	Obere Grenze
Mittelwert	Starke Anstrengung (pro Woche)	3,73	,033	3,66	3,79
	Moderate Anstrengung (pro Woche)	4,90	,041	4,82	4,98
	Krafttraining (pro Woche)	3,52	,042	3,43	3,60

Jede der ausgewählten Statistiken wird für jede der Maßvariablen berechnet. Die erste Spalte enthält Schätzwerte für die durchschnittliche Häufigkeit pro Woche, die sich eine Person einer bestimmten Art von Aktivität widmet. Die Konfidenzintervalle für die Mittelwerte überschneiden sich nicht. Daraus lässt sich ableiten, dass die Amerikaner sich im Allgemeinen weniger häufig dem Kraftraining widmen als intensiver körperlicher Bewegung und weniger häufig intensiver körperlicher Bewegung als mäßiger körperlicher Bewegung.

Univariate Statistiken nach Teilgesamtheit

Abbildung 16-5
Univariate Statistiken nach Teilgesamtheit

Altersgruppe			Schätzung	Standard-fehler	95%-Konfidenzintervall	
					Untere Grenze	Obere Grenze
18-24	Mittelwert	Starke Anstrengung (pro Woche)	3,92	,087	3,75	4,09
		Moderate Anstrengung (pro Woche)	5,18	,137	4,91	5,45
		Krafttraining (pro Woche)	3,45	,085	3,28	3,62
25-44	Mittelwert	Starke Anstrengung (pro Woche)	3,55	,048	3,46	3,65
		Moderate Anstrengung (pro Woche)	4,73	,056	4,62	4,84
		Krafttraining (pro Woche)	3,28	,052	3,18	3,38
45-64	Mittelwert	Starke Anstrengung (pro Woche)	3,79	,063	3,66	3,91
		Moderate Anstrengung (pro Woche)	4,88	,070	4,74	5,02
		Krafttraining (pro Woche)	3,65	,092	3,47	3,84
65+	Mittelwert	Starke Anstrengung (pro Woche)	4,18	,111	3,96	4,39
		Moderate Anstrengung (pro Woche)	5,22	,084	5,06	5,39
		Krafttraining (pro Woche)	4,66	,155	4,36	4,97

Jede ausgewählte Statistik wird für jede Maßvariable nach den Werten der *Age category* (Altersgruppe) berechnet. Die erste Spalte enthält Schätzwerte für die durchschnittliche Häufigkeit pro Woche, die sich die Personen in den einzelnen Kategorien einer bestimmten Art

von Aktivität widmen. Aus den Konfidenzintervallen für die Mittelwerte können Sie interessante Schlussfolgerungen ziehen.

- Was die intensive und mäßige körperliche Bewegung betrifft, sind die 25–44-Jährigen weniger aktiv als die 18–24-Jährigen und die 45–64-Jährigen. Außerdem sind die 45–64-Jährigen weniger aktiv als die Altersgruppe der mindestens 65-Jährigen.
- Was das Krafttraining betrifft, sind die 25–44-Jährigen weniger aktiv als die 45–64. Außerdem sind die 18–24-Jährigen und die 45–64-Jährigen weniger aktiv als die Altersgruppe der mindestens 65-Jährigen.

Zusammenfassung

Mit der Prozedur "Deskriptive Statistiken für komplexe Stichproben" haben Sie Statistiken für die Aktivitätsniveaus von US-Bürgern erhalten.

- Insgesamt wenden die Amerikaner unterschiedlich viel Zeit für unterschiedliche Arten von sportlicher Aktivität auf.
- Aufgeschlüsselt nach Alter ergibt sich grob gesagt, dass Amerikaner nach dem College-Alter zunächst weniger aktiv sind als in der Schule, jedoch mit steigendem Alter körperliche Betätigung wieder ernster nehmen.

Verwandte Prozeduren

Die Prozedur "Deskriptive Statistiken für komplexe Stichproben" ist ein nützliches Werkzeug zur Gewinnung deskriptiver Statistiken von Skalenmaßen für Beobachtungen, die mittels eines komplexen Stichprobenplans gewonnen wurden.

- Der Stichprobenassistent für komplexe Stichproben wird zur Angabe der Planspezifikationen für komplexe Stichproben und zum Ziehen von Stichproben verwendet. Die vom Stichprobenassistenten erstellte Stichprobenplan-Datei enthält einen Standard-Analyseplan und kann im Dialogfeld "Plan" angegeben werden, wenn die gezogene Stichprobe gemäß diesem Plan analysiert werden soll.
- Der Analysevorbereitungsassistent für komplexe Stichproben wird zur Angabe der Analysespezifikationen für eine bestehende komplexe Stichprobe verwendet. Die vom Stichprobenassistenten erstellte Analyseplan-Datei kann im Dialogfeld "Plan" angegeben werden, wenn Sie die Stichprobe gemäß diesem Plan analysieren.
- Die Prozedur Verhältnisse für komplexe Stichproben bietet deskriptive Statistiken für Verhältnisse von Skalenmaßen.
- Die Prozedur Häufigkeiten für komplexe Stichprobe bietet univariate deskriptive Statistiken für kategoriale Variablen.

Kapitel
17

Kreuztabellen für komplexe Stichproben

Mit der Prozedur “Kreuztabellen für komplexe Stichproben” werden Kreuztabellen für Paare von ausgewählten Variablen erstellt und bivariate Statistiken angezeigt. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Verwendung von “Kreuztabellen für komplexe Stichproben” zum Messen des relativen Risikos eines Ereignisses

Ein Unternehmen, das Zeitschriftenabonnements vertreibt, sendet üblicherweise jeden Monat Post an Personen aus einer gekauften Namensdatenbank. Die Antwortrate ist normalerweise gering, sodass nach besseren Methoden gesucht werden muss, um potenzielle Kunden anzusprechen. Ein Vorschlag besteht darin, die Postsendungen auf Personen mit Zeitungsabonnements zu konzentrieren, da anzunehmen ist, dass Personen, die Zeitungen lesen, mit größerer Wahrscheinlichkeit ein Zeitschriftenabonnement abschließen.

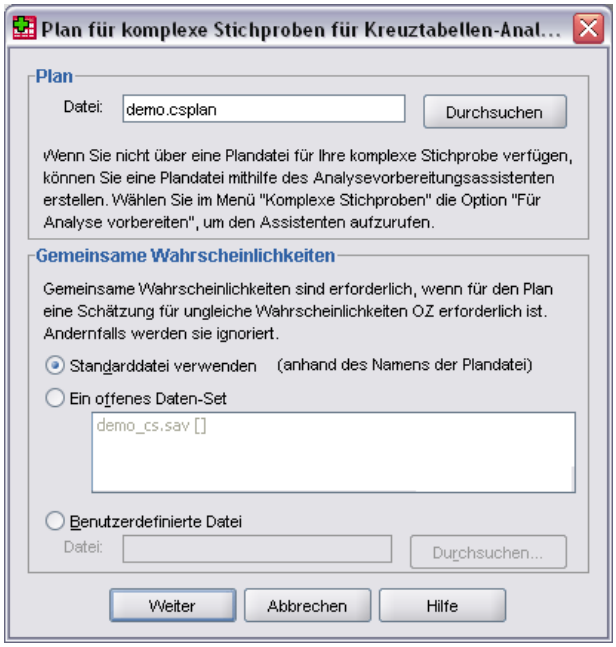
Verwenden Sie die Prozedur “Kreuztabellen für komplexe Stichproben”, um diese Theorie zu testen, indem Sie eine 2x2-Tabelle (*Zeitungsabonnement* zu *Antwort*) erstellen und das relative Risiko berechnen, mit dem eine Person mit dem Abschluss eines Zeitungsabonnements auf die Postsendung reagiert. Diese Informationen finden Sie in der Datei *demo_cs.sav*, die mit der Stichprobenplan-Datei *demo.csplan* analysiert werden sollte. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271.

Durchführung der Analyse

- Um eine Analyse der Art “Kreuztabellen für komplexe Stichproben” durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

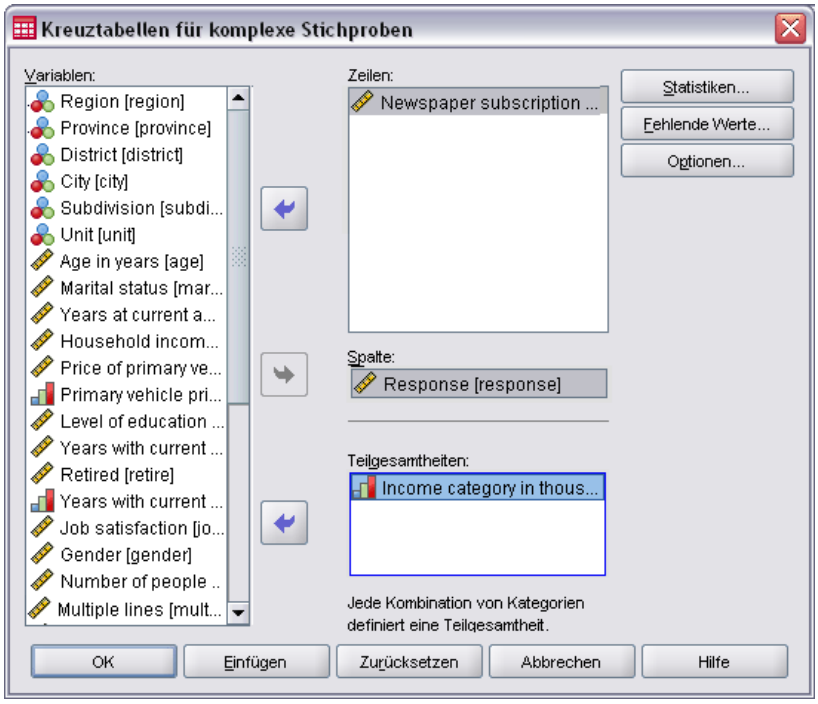
  Analysieren
  Komplexe Stichproben
  Kreuztabellen...

Abbildung 17-1
Dialogfeld "Komplexe Stichproben: Plan"


Plan für komplexe Stichproben für Kreuztabellen-Anal...
Plan
Datei:
demo.csplan
Durchsuchen
Wenn Sie nicht über eine Plandatei für Ihre komplexe Stichprobe verfügen, können Sie eine Plandatei mithilfe des Analysevorbereitungsassistenten erstellen. Wählen Sie im Menü "Komplexe Stichproben" die Option "Für Analyse vorbereiten", um den Assistenten aufzurufen.
Gemeinsame Wahrscheinlichkeiten
Gemeinsame Wahrscheinlichkeiten sind erforderlich, wenn für den Plan eine Schätzung für ungleiche Wahrscheinlichkeiten OZ erforderlich ist. Andernfalls werden sie ignoriert.
Standarddatei verwenden (anhand des Namens der Plandatei)
Ein offenes Daten-Set
demo_cs.sav []
Benutzerdefinierte Datei
Datei:
Durchsuchen...
Weiter
Abbrechen
Hilfe

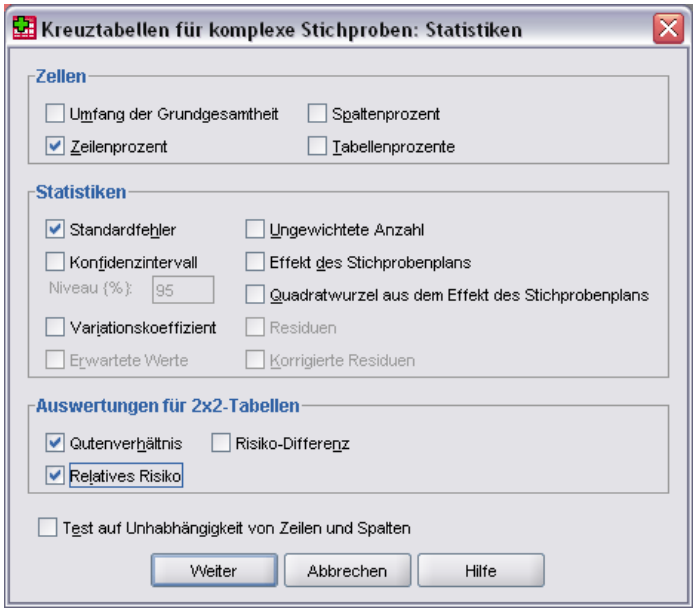
- Wechseln Sie zu der Datei *demo.csplan* und wählen Sie sie aus. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271.
- Klicken Sie auf Weiter.

Abbildung 17-2
Dialogfeld "Kreuztabellen"



- Wählen Sie *Newspaper subscription* (Zeitungsabonnement) als Zeilenvariable.
- Wählen Sie *Response* (Antwort) als Spaltenvariable.
- Da die Aufschlüsselung der Ergebnisse nach Einkommensklassen ebenfalls interessant ist, wählen Sie *Income category in thousands* (Einkommensklasse in Tausend) als Teilgesamtheitsvariable aus.
- Klicken Sie auf Statistik.

Abbildung 17-3
Dialogfeld "Kreuztabellen: Statistiken"



- Heben Sie die Auswahl von Population size (Umfang der Grundgesamtheit) auf und wählen Sie im Gruppenfeld "Zellen" Row percent (Zeilenprozentsatz) aus.
- Wählen Sie Odds ratio (Quotenverhältnis) und Relative risk (Relatives Risiko) in der Auswertung für das Gruppenfeld der 2x2-Tabelle aus.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Kreuztabellen für komplexe Stichproben" auf OK.

  Durch diese Auswahl wird eine Kreuztabelle und ein Risikoschätzer für *Newspaper subscription* (Zeitungsabonnement) zu *Response* (Antwort) erstellt. Außerdem werden separate Tabellen erstellt, in denen die Ergebnisse nach *Income category in thousands* (Einkommensklasse in Tausend) aufgeschlüsselt sind.

Kreuztabellen

Abbildung 17-4
Kreuztabelle für "Zeitungsabonnement" und "Antwort"

Abonnement Tageszeitung			Antwort		
			Ja	Nein	Insgesamt
Ja	% von Abonnement Tageszeitung	Schätzung	17,2%	82,8%	100,0%
		Standardfehler	1,0%	1,0%	,0%
Nein	% von Abonnement Tageszeitung	Schätzung	10,3%	89,7%	100,0%
		Standardfehler	,7%	,7%	,0%
Insgesamt	% von Abonnement Tageszeitung	Schätzung	12,8%	87,2%	100,0%
		Standardfehler	,7%	,7%	,0%

Die Kreuztabelle zeigt, dass insgesamt recht wenige Personen auf die Postsendung reagiert haben. Bei den Zeitungsabonnenten lag der Anteil der Antworten jedoch höher.

Risikoschätzer

Abbildung 17-5
Risikoschätzer für "Zeitungsabonnement" und "Antwort"

			Schätzung
Abonnement Tageszeitung * Antwort	Quotenverhältnis		1,812
	Relatives Risiko	Für Kohorte Antwort = Ja	1,673
		Für Kohorte Antwort = Nein	,923

Statistiken werden nur für 2x2-Tabellen unter Berücksichtigung aller Zellen berechnet.

Das relative Risiko ist ein Quotient aus Ereigniswahrscheinlichkeiten. Das relative Risiko für eine Antwort auf die Postsendung ist der Quotient aus der Wahrscheinlichkeit, dass ein Zeitungsabonnent antwortet, und der Wahrscheinlichkeit, dass ein Nicht-Abonnent antwortet. Daher ist der Schätzer für das relative Risiko einfach 17,2 % : 10,3% = 1,673. Entsprechend ist das relative Risiko für die Nichtantwort der Quotient aus der Wahrscheinlichkeit, dass ein Abonnent nicht antwortet, und der Wahrscheinlichkeit, dass ein Nicht-Abonnent nicht antwortet. Ihr Schätzer für dieses relative Risiko ist 0,923. Anhand dieser Ergebnisse können Sie abschätzen, dass ein Zeitungsabonnent gegenüber einem Nicht-Abonnenten mit 1,673-facher Wahrscheinlichkeit auf die Postsendung antwortet bzw. dass er gegenüber einem Nicht-Abonnenten mit 0,923-facher Wahrscheinlichkeit nicht antwortet.

Das Quotenverhältnis ist ein Quotient der Ereignschancen. Die Chance für ein Ereignis ist der Quotient aus der Wahrscheinlichkeit, dass das Ereignis eintritt, und der Wahrscheinlichkeit, dass das Ereignis nicht eintritt. Daher ist der Schätzer für die Chance, dass ein Abonnent auf die Postsendung reagiert 17,2% : 82,8% = 0,208. Entsprechend ist der Schätzer für die Chance, dass ein Nicht-Abonnent auf die Postsendung reagiert 10,3% : 89,7% = 0,115. Der Schätzer für das Quotenverhältnis ist daher 0,208 : 0,115 = 1,812 (beachten Sie, dass es in den Zwischenschritten zu einem gewissen Rundungsfehler kommt). Das Quotenverhältnis ist auch der Quotient aus dem relativen Antwortrisiko und dem relativen Risiko, dass keine Antwort erfolgt, also 1,673 : 0,923 = 1,812.

Quotenverhältnis im Vergleich zum relativen Risiko

Da es sich dabei um einen Quotient aus Quotienten handelt, ist das Quotenverhältnis sehr schwer zu interpretieren. Die Interpretation des relativen Risikos ist einfacher, daher ist das Quotenverhältnis allein nicht sehr hilfreich. Es gibt jedoch bestimmte, häufig vorkommende Situationen, bei denen der Schätzer für das relative Risiko nicht besonders gut ist und in denen das Quotenverhältnis verwendet werden kann, um das relative Risiko für das untersuchte Ereignis näherungsweise abzuschätzen. Das Quotenverhältnis sollte als Approximation für das relative

Risiko des untersuchten Ereignisses verwendet werden, wenn beide der folgenden Bedingungen vorliegen:

- Die Wahrscheinlichkeit für das untersuchte Ereignis ist niedrig (<0,1). Diese Bedingung garantiert, dass das Quotenverhältnis eine gute Approximation für das relative Risiko darstellt. In diesem Beispiel ist das untersuchte Ereignis eine Antwort auf die Postsendung.
- Bei der Studie handelt es sich um eine Fall-Kontroll-Studie. Diese Bedingung bedeutet, dass der übliche Schätzer für das relative Risiko mit hoher Wahrscheinlichkeit nicht gut ist. Eine Fall-Kontroll-Studie ist retrospektiv und wird besonders häufig verwendet, wenn das untersuchte Ereignis unwahrscheinlich ist oder wenn ein prospektiver Experimentaufbau aus praktischen oder ethischen Gründen nicht infrage kommt.

Im vorliegenden Beispiel ist keine der genannten Bedingungen erfüllt, da der Gesamtprozentsatz der antwortenden Personen 12,8 betrug und es sich nicht um eine Fall-Kontroll-Studie handelte. Daher ist es sicherer, 1,673 als relatives Risiko anzugeben als den Wert des Quotenverhältnisses.

Risikoschätzer nach Teilgesamtheit

Abbildung 17-6
Risikoschätzer für "Zeitungsabonnement" und "Antwort," nach Einkommensklasse kontrolliert.

Einkommensgruppen				Schätzung
Under€25	Abonnement Tageszeitung * Antwort	Quotenverhältnis		2.712
		Relatives Risiko	Für Kohorte Antwort = Ja	2.241
			Für Kohorte Antwort = Nein	.826
€25 - €49	Abonnement Tageszeitung * Antwort	Quotenverhältnis		1.794
		Relatives Risiko	Für Kohorte Antwort = Ja	1.645
			Für Kohorte Antwort = Nein	.917
€50 - €74	Abonnement Tageszeitung * Antwort	Quotenverhältnis		1.168
		Relatives Risiko	Für Kohorte Antwort = Ja	1.152
			Für Kohorte Antwort = Nein	.986
€75+	Abonnement Tageszeitung * Antwort	Quotenverhältnis		1.242
		Relatives Risiko	Für Kohorte Antwort = Ja	1.227
			Für Kohorte Antwort = Nein	.988

Statistiken werden nur für 2x2-Tabellen unter Berücksichtigung aller Zellen berechnet.

Die relativen Risikoschätzer werden für jede Einkommensklasse getrennt berechnet. Beachten Sie: Das relative Risiko einer positiven Antwort scheint bei den Zeitungsabonnenten mit zunehmendem Einkommen nach und nach abzunehmen, was darauf schließen lässt, dass Sie die Zielgruppe für die Postsendungen eventuell noch weiter eingrenzen können.

Zusammenfassung

Es ergab sich, dass mithilfe der Risikoschätzer aus "Kreuztabellen für komplexe Stichproben" die Antwortrate auf Postsendungen erhöht werden kann, indem die Zielgruppe auf Zeitungsabonnenten verkleinert wird. Des Weiteren deutete einiges darauf hin, dass die Risikoschätzer nicht für alle *Einkommensklassen* konstant sind, sodass die Antwortrate durch

Eingrenzen der Zielgruppe auf Zeitungsabonnenten mit relativ niedrigem Einkommen eventuell noch weiter gesteigert werden kann.

Verwandte Prozeduren

Die Prozedur “Kreuztabellen für komplexe Stichproben” ist ein nützliches Werkzeug zur Gewinnung deskriptiver Statistiken aus Kreuztabellen von kategorialen Variablen für Beobachtungen, die mittels eines komplexen Stichprobenplans gewonnen wurden.

- Der Stichprobenassistent für komplexe Stichproben wird zur Angabe der Planspezifikationen für komplexe Stichproben und zum Ziehen von Stichproben verwendet. Die vom Stichprobenassistenten erstellte Stichprobenplan-Datei enthält einen Standard-Analyseplan und kann im Dialogfeld “Plan” angegeben werden, wenn die gezogene Stichprobe gemäß diesem Plan analysiert werden soll.
- Der Analysevorbereitungsassistent für komplexe Stichproben wird zur Angabe der Analysespezifikationen für eine bestehende komplexe Stichprobe verwendet. Die vom Stichprobenassistenten erstellte Analyseplan-Datei kann im Dialogfeld “Plan” angegeben werden, wenn Sie die Stichprobe gemäß diesem Plan analysieren.
- Die Prozedur Häufigkeiten für komplexe Stichprobe bietet univariate deskriptive Statistiken für kategoriale Variablen.

Kapitel
18

Verhältnisse für komplexe Stichproben

Die Prozedur “Verhältnisse für komplexe Stichproben” zeigt univariate Auswertungsstatistiken für Verhältnisse von Variablen an. Optional können Sie Statistiken nach Untergruppen anfordern, die durch eine oder mehrere kategoriale Variablen definiert sind.

Verwenden von “Verhältnisse für komplexe Stichproben” zur Erleichterung der Schätzung von Immobilienwerten

Eine bundesstaatliche Behörde ist damit beauftragt zu gewährleisten, dass die Vermögenssteuer in den einzelnen Counties gerecht bemessen wird. Die Steuern beruhen auf der Schätzung des Immobilienwerts. Daher möchte die Behörde die Immobilienwerte in allen Counties untersuchen, um sicherzugehen, dass die Akten jedes County gleichermaßen auf dem neuesten Stand sind. Da die Ressourcen für die Gewinnung aktueller Schätzungen begrenzt sind, entschied sich die Behörde für die Anwendung eines Verfahrens mit komplexen Stichproben zur Auswahl von Immobilien.

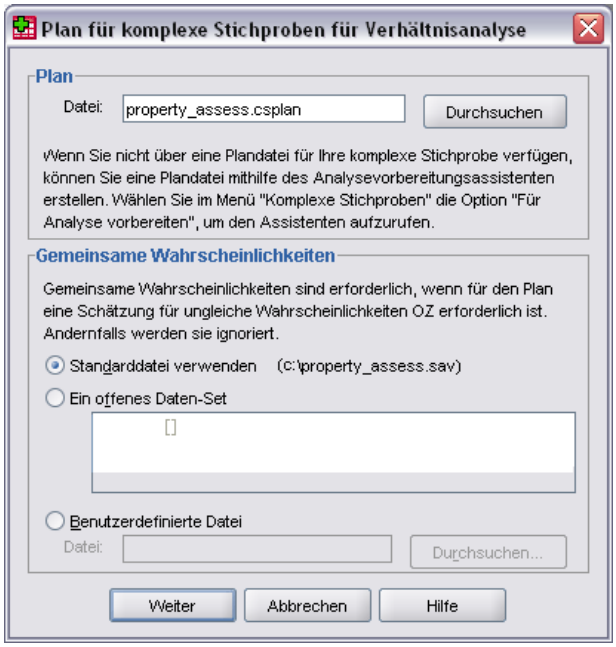
Die Stichprobe der ausgewählten Immobilien und der zugehörigen Informationen zur Schätzung ihres Werts finden Sie in *property_assess_cs_sample.sav*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Verwenden Sie “Verhältnisse für komplexe Stichproben”, um die Veränderung in den Immobilienwerten in allen fünf Counties seit der letzten Schätzung zu bewerten.

Durchführung der Analyse

▶ Um eine Analyse der Art “Verhältnisse für komplexe Stichproben” durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

Analysieren
  Komplexe Stichproben
    Verhältnisse...

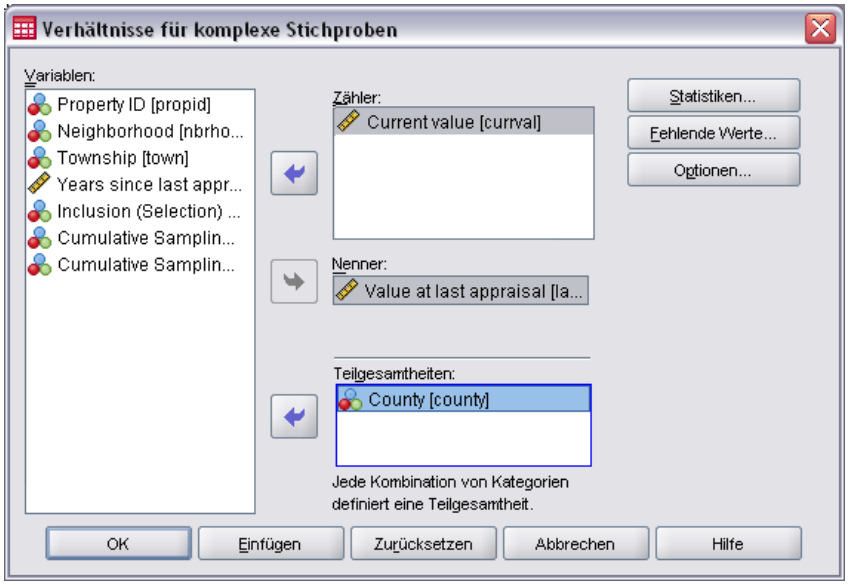
Abbildung 18-1
Dialogfeld "Komplexe Stichproben: Plan"


Plan für komplexe Stichproben für Verhältnisanalyse
Plan
Datei:
property_assess.csplan
Durchsuchen
Wenn Sie nicht über eine Plandatei für Ihre komplexe Stichprobe verfügen, können Sie eine Plandatei mithilfe des Analysevorbereitungsassistenten erstellen. Wählen Sie im Menü "Komplexe Stichproben" die Option "Für Analyse vorbereiten", um den Assistenten aufzurufen.
Gemeinsame Wahrscheinlichkeiten
Gemeinsame Wahrscheinlichkeiten sind erforderlich, wenn für den Plan eine Schätzung für ungleiche Wahrscheinlichkeiten OZ erforderlich ist. Andernfalls werden sie ignoriert.
Standarddatei verwenden (c:\property_assess.sav)
Ein offenes Daten-Set
[]
Benutzerdefinierte Datei
Datei:
Durchsuchen...
Weiter
Abbrechen
Hilfe

- Wechseln Sie zu der Datei *property_assess.csplan* und wählen Sie sie aus. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271.

- Klicken Sie auf Weiter.

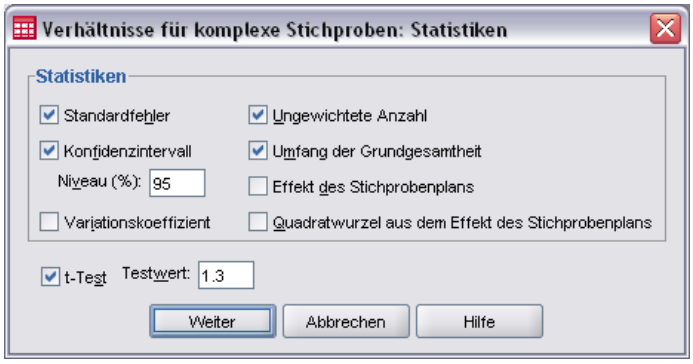
Abbildung 18-2
Dialogfeld "Verhältnisse"



- Wählen Sie *Current value* (Aktueller Wert) als Zählervariable aus.

- Wählen Sie *Value at last appraisal* (Wert bei der letzten Schätzung) als Nennervariable aus.
- Wählen Sie *County* als Teilgesamtheitsvariable aus.
- Klicken Sie auf Statistik.

Abbildung 18-3
Dialogfeld "Verhältnisse: Statistiken"



- Wählen Sie im Gruppenfeld "Statistiken" die Optionen Konfidenzintervall, Ungewichtete Anzahl und Umfang der Grundgesamtheit aus.
- Wählen Sie T-Test aus und geben Sie als Testwert 1,3 ein.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Verhältnisse für komplexe Stichproben" auf OK.

Verhältnisse

Abbildung 18-4
Verhältnistabelle

County	Zähler	Nenner	Verhältnisschätzung	Standardfehler	95%-Konfidenzintervall	
					Untere	Obere
Osten	Current value	Value at last appraisal	1,381	,068	1,236	1,525
Mitte	Current value	Value at last appraisal	1,364	,064	1,227	1,502
Westen	Current value	Value at last appraisal	1,524	,053	1,410	1,638
Norden	Current value	Value at last appraisal	1,277	,032	1,208	1,346
Süden	Current value	Value at last appraisal	1,195	,029	1,134	1,256

Die Standardansicht der Tabelle ist sehr breit. Daher müssen Sie zur besseren Übersicht pivotieren.

Pivotieren der Verhältnistabelle

- Doppelklicken Sie auf die Tabelle, um sie zu aktivieren.
- Wählen Sie die folgenden Befehle aus den Menüs des Viewers aus:

Pivot
Pivot-Leisten

- Ziehen Sie *Nenner* und anschließend *Zähler* aus der Zeile in die Schicht.

- ▶ Ziehen Sie *County* aus der Zeile in die Spalte.
- ▶ Ziehen Sie *Statistiken* aus der Spalte in die Zeile.
- ▶ Schließen Sie das Fenster "Pivot-Leisten".

Pivotierte Verhältnistabelle

Abbildung 18-5
Pivotierte Verhältnistabelle

Nenner Current value
Zähler Value at last appraisal

		Region				
		Osten	Mitte	Westen	Norden	Süden
Verhältnisschätzung		1.381	1.364	1.524	1.277	1.195
Standardfehler		.068	.064	.053	.032	.029
95%-Konfidenzintervall	Untere Grenze	1.236	1.227	1.410	1.208	1.134
	Obere Grenze	1.525	1.502	1.638	1.346	1.256
Testen von Hypothesen	Testwert	1.3	1.3	1.3	1.3	1.3
	t	1.191	.997	4.201	-.702	-3.646
	Freiheitsgrade	15	15	15	15	15
	Sig.	.252	.334	.001	.493	.002
Ungewichtete Anzahl		168	179	202	205	220

Die Verhältnistabelle ist nun pivotiert, sodass die Statistiken sich leichter zwischen den verschiedenen Counties vergleichen lassen.

- Die Verhältnisschätzungen reichen vom niedrigsten Wert von 1,195 in "Southern County" bis zu einem Höchstwert von 1,524 in "Western County".
- Außerdem besteht eine beträchtliche Variabilität in den Standardfehlern, die vom niedrigsten Wert von 0,029 in "Southern County" bis zu einem Höchstwert von "0,068" in "Eastern County" reichen.
- Einige der Konfidenzintervalle überschneiden sich nicht. Daraus lässt sich ableiten, dass die Verhältnisse für "Western County" höher liegen als die Verhältnisse für "Northern County" und "Southern County".
- Schließlich liegen die Signifikanzwerte (ein objektiveres Maß) für die *T*-Tests für "Western County" und "Southern County" unter 0,05. Daraus lässt sich ableiten, dass das Verhältnis für "Western County" über 1,3 und für "Southern County" unter 1,3 liegt.

Zusammenfassung

Mithilfe der Prozedur "Verhältnisse für komplexe Stichproben" haben Sie verschiedene Statistiken für das Verhältnis zwischen *Current value* (Aktueller Wert) und *Value at last appraisal* (Wert bei der letzten Schätzung) gewonnen. Die Ergebnisse legen nahe, dass gewisse Ungleichheiten in

der Bemessung der Vermögenssteuern zwischen den einzelnen Counties vorliegen könnten. Insbesondere handelt es sich dabei um Folgendes:

- Die Verhältnisse für "Western County" sind hoch, was darauf hindeutet, dass die dortigen Akten in Bezug auf die Bewertung von Immobilienwerten nicht so aktuell sind wie die in den anderen Counties. Die Vermögenssteuern in diesem County sind vermutlich zu niedrig.
- Die Verhältnisse für "Southern County" sind niedrig, was darauf hindeutet, dass die dortigen Akten in Bezug auf die Bewertung von Immobilienwerten aktueller sind als die in den anderen Counties. Die Vermögenssteuern in diesem County sind vermutlich zu hoch.
- Die Verhältnisse für "Southern County" sind niedriger als die für "Western County", liegen jedoch noch immer im Zielbereich von 1,3.

Die für die Untersuchung von Immobilienwerten in "Southern County" eingesetzten Ressourcen werden nun "Western County" zugewiesen, um die Verhältnisse für diese Counties mit den Verhältnissen für die anderen Counties und dem Ziel von 1,3 in Einklang zu bringen.

Verwandte Prozeduren

Die Prozedur "Verhältnisse für komplexe Stichproben" ist ein nützliches Werkzeug zur Gewinnung deskriptiver Statistiken für das Verhältnis von Skalenmaßen für Beobachtungen, die mittels eines komplexen Stichprobenplans gewonnen wurden.

- Der Stichprobenassistent für komplexe Stichproben wird zur Angabe der Planspezifikationen für komplexe Stichproben und zum Ziehen von Stichproben verwendet. Die vom Stichprobenassistenten erstellte Stichprobenplan-Datei enthält einen Standard-Analyseplan und kann im Dialogfeld "Plan" angegeben werden, wenn die gezogene Stichprobe gemäß diesem Plan analysiert werden soll.
- Der Analysevorbereitungsassistent für komplexe Stichproben wird zur Angabe der Analysespezifikationen für eine bestehende komplexe Stichprobe verwendet. Die vom Stichprobenassistenten erstellte Analyseplan-Datei kann im Dialogfeld "Plan" angegeben werden, wenn Sie die Stichprobe gemäß diesem Plan analysieren.
- Die Prozedur Deskriptive Statistiken für komplexe Stichproben bietet deskriptive Statistiken für metrische Variablen.

Kapitel
19

Allgemeines lineares Modell für komplexe Stichproben

Die Prozedur "Allgemeines lineares Modell für komplexe Stichproben" besteht aus einer linearen Regressionsanalyse sowie aus einer Analyse der Varianz und Kovarianz für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Optional können Sie auch Analysen für eine Teilgesamtheit vornehmen.

Verwendung des allgemeinen linearen Modells für komplexe Stichproben zur Anpassung einer zweifaktoriellen ANOVA

Eine Lebensmittelkette hat eine Kundenumfrage über die Kaufgewohnheiten durchgeführt, die nach einem komplexen Plan ausgeführt wurde. Auf der Grundlage der Umfrageergebnisse und der Zahlen über die Ausgaben der einzelnen Kunden im vergangenen Monat möchte das Unternehmen ermitteln, ob die Einkaufshäufigkeit in einem Zusammenhang mit den monatlichen Ausgaben steht, und zwar getrennt nach Geschlecht. Bei dieser Untersuchung soll der Stichprobenplan berücksichtigt werden.

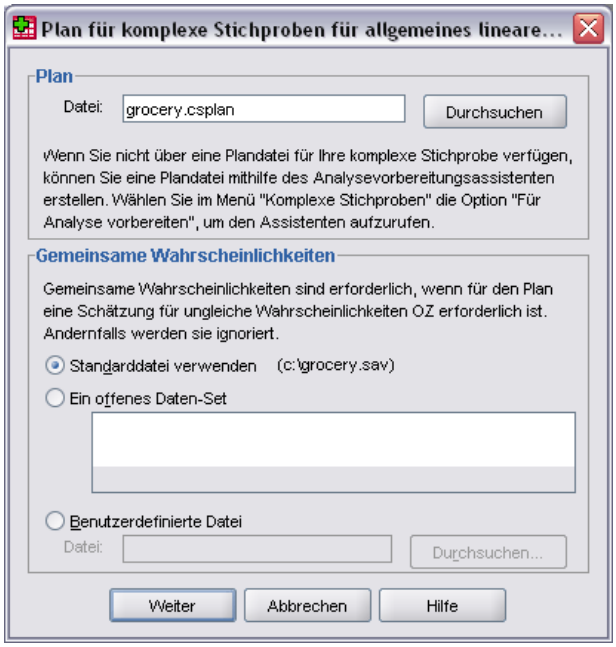
Diese Informationen finden Sie in der Datei *grocery_1month_sample.sav*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Mit der Prozedur "Allgemeines lineares Modell für komplexe Stichproben" können Sie eine zweifaktorielle ANOVA (Zweifach-Anova) für den ausgegebenen Betrag durchführen.

Durchführung der Analyse

- Um eine Analyse der Art "Allgemeines lineares Modell für komplexe Stichproben" durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

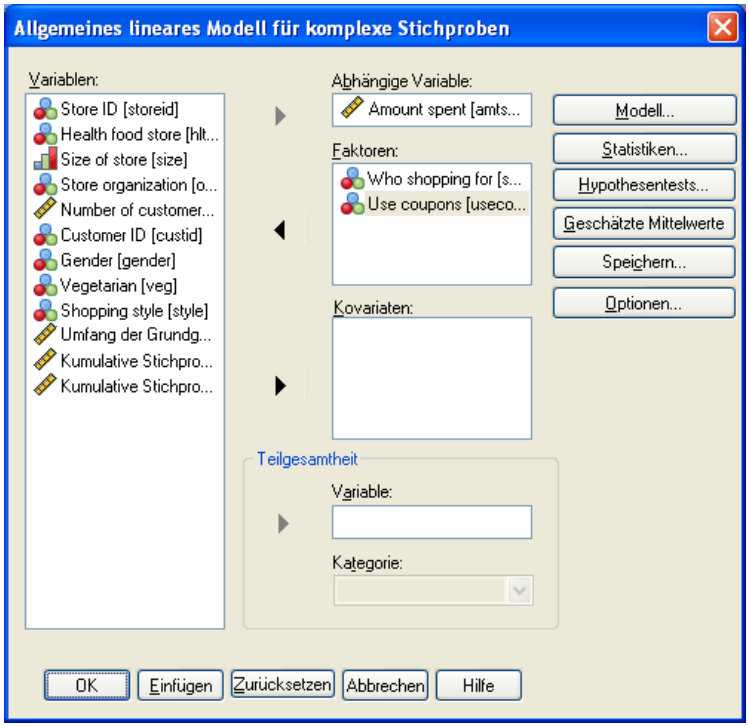
Analysieren
  Komplexe Stichproben
    Allgemeines lineares Modell...

Abbildung 19-1
Dialogfeld "Komplexe Stichproben: Plan"


Plan für komplexe Stichproben für allgemeines lineare...
Plan
Datei:
grocery.csplan
Durchsuchen
Wenn Sie nicht über eine Plandatei für Ihre komplexe Stichprobe verfügen, können Sie eine Plandatei mithilfe des Analysevorbereitungsassistenten erstellen. Wählen Sie im Menü "Komplexe Stichproben" die Option "Für Analyse vorbereiten", um den Assistenten aufzurufen.
Gemeinsame Wahrscheinlichkeiten
Gemeinsame Wahrscheinlichkeiten sind erforderlich, wenn für den Plan eine Schätzung für ungleiche Wahrscheinlichkeiten OZ erforderlich ist. Andernfalls werden sie ignoriert.
Standarddatei verwenden (c:\grocery.sav)
Ein offenes Daten-Set
Benutzerdefinierte Datei
Datei:
Durchsuchen...
Weiter
Abbrechen
Hilfe

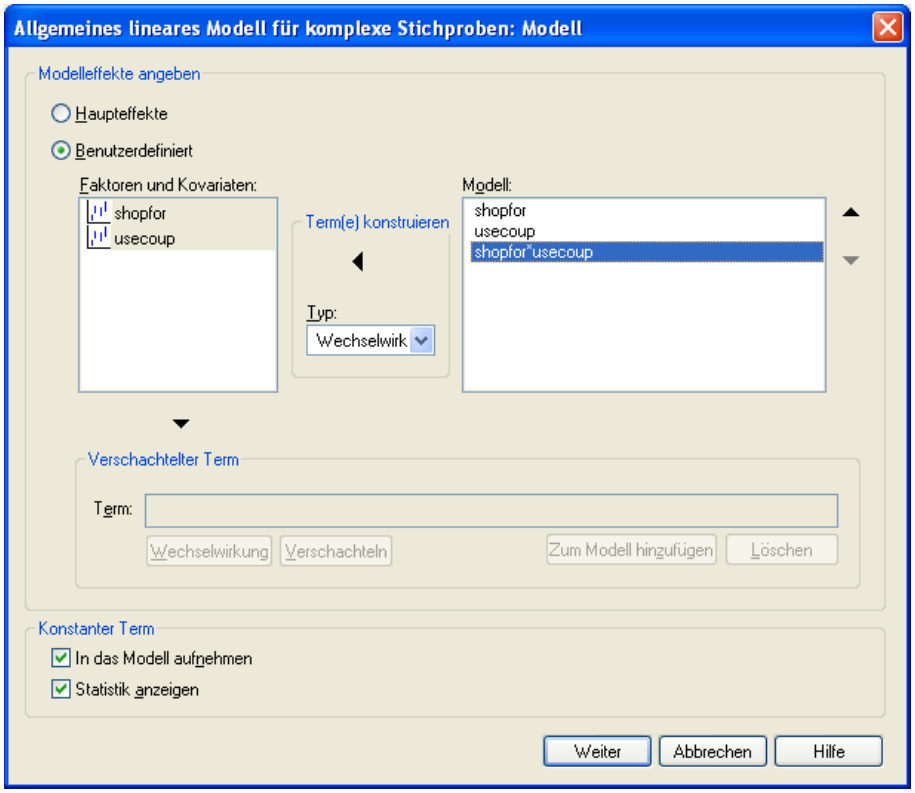
- Wechseln Sie zu der Datei *grocery.csplan* und wählen Sie sie aus. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271.
- Klicken Sie auf **Weiter**.

Abbildung 19-2
Dialogfeld "Allgemeines lineares Modell"



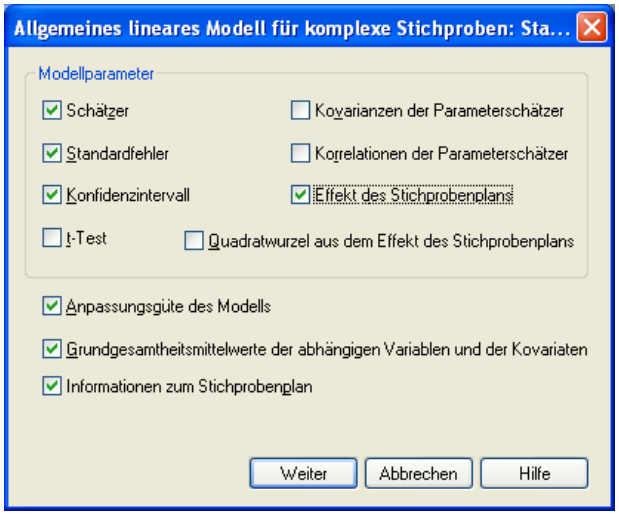
- Wählen Sie *Amount spent* (Ausgegebener Betrag) als abhängige Variable aus.
- Wählen Sie *Who shopping for* (Einkauf für wen) und *Use coupons* (Verwendung von Coupons) als Faktoren aus.
- Klicken Sie auf Modell.

Abbildung 19-3
Dialogfeld "Modell"



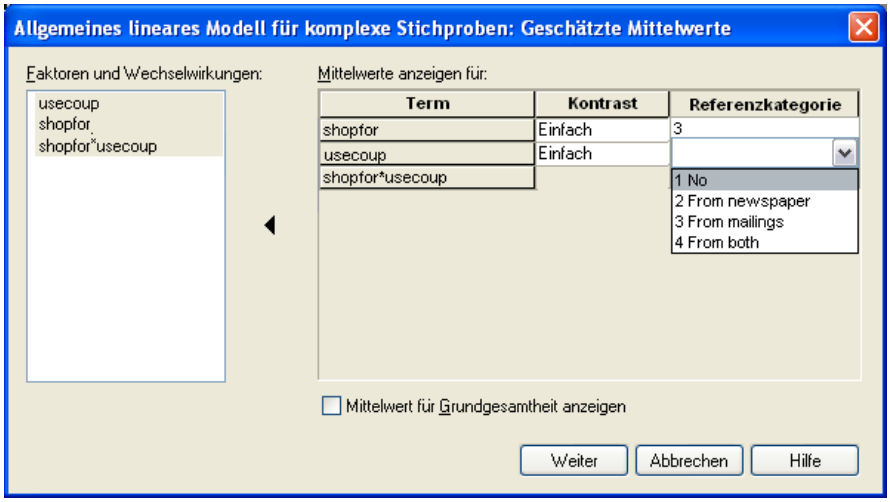
- Wählen Sie aus, dass ein benutzerdefiniertes Modell erstellt werden soll.
- Wählen Sie Haupteffekte als Typ für den zu erstellenden Term aus und wählen Sie *shopfor* und *usecoup* als Modellterme aus.
- Wählen Sie Wechselwirkung als Typ für den zu erstellenden Term aus und fügen Sie die Wechselwirkung *shopfor*usecoup* als Modellterm hinzu.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Allgemeines lineares Modell" auf Statistik.

Abbildung 19-4
Dialogfeld "Allgemeines lineares Modell: Statistik"



- Wählen Sie Schätzer, Standardfehler, Konfidenzintervall und Effekt des Stichprobenplans in der Gruppe "Modellparameter" aus.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Allgemeines lineares Modell" auf Geschätzte Mittelwerte.

Abbildung 19-5
Dialogfeld "Allgemeines lineares Modell: Geschätzte Mittelwerte"



- Wählen Sie diese Option, um Mittelwerte für *shopfor*, *usecoup* und die Wechselwirkung *shopfor*usecoup* anzuzeigen.
- Wählen Sie Einfach als Kontrast und 3 Self and family (3 Eigene Person und Familie) als Referenzkategorie für *shopfor*. Beachten Sie: Nach der Auswahl wird für die Kategorie "3" im Dialogfeld angezeigt.
- Wählen Sie Einfach als Kontrast und 1 No (1 Nein) als Referenzkategorie für *usecoup*.

- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Allgemeines lineares Modell" auf OK.

Modellzusammenfassung

Abbildung 19-6
R-Quadrat-Statistik

R-Quadrat	,601

a. Modell: Amount spent = (Konstanter Term) + shopfor + usecoup + shopfor * usecoup

R-Quadrat, das Bestimmtheitsmaß, ist ein Maß für die Anpassungsgüte des Modells. Es zeigt sich, dass ca. 60 % der Schwankungen bei *Amount spent* (Ausgegebener Betrag) durch das Modell erklärt werden, was eine gute Erklärungsleistung darstellt. Dennoch können Sie weitere Einflussvariablen in das Modell aufnehmen, um die Anpassung weiter zu verbessern.

Tests der Modelleffekte

Abbildung 19-7
Tests der Zwischensubjekteffekte

Quelle	Freiheits-grade 1	Freiheits-grade 2	Wald-F	Sig.
(Korrigiertes Modell)	11,000	3,000	127,231	,001
(Konstanter Term)	1,000	13,000	6321,597	,000
shopfor	2,000	12,000	643,593	,000
usecoup	3,000	11,000	87,453	,000
shopfor * usecoup	6,000	8,000	10,688	,002

a. Modell: Amount spent = (Konstanter Term) + shopfor + usecoup + shopfor * usecoup

Jeder Term im Modell sowie das Modell als Ganzes werden daraufhin getestet, ob der Wert seines Effekts gleich 0 ist. Terme mit Signifikanzwerten von weniger als 0,05 weisen einen erkennbaren Effekt auf. Alle Modellterme tragen also zum Modell bei.

Parameterschätzer

Abbildung 19-8
Parameterschätzer

Parameter	Schätzung	Standard-fehler	95%-Konfidenzintervall		Effekt des Stich-probenplans
			Untere Grenze	Obere Grenze	
(Konstanter Term)	518,249	11,731	492,905	543,592	1,387
[shopfor=1]	-174,757	10,762	-198,006	-151,508	,950
[shopfor=2]	-129,443	11,455	-154,191	-104,696	,925
[shopfor=3]	,000[a]	.	.	.	.
[usecoup=1]	-140,838	10,180	-162,830	-118,846	,649
[usecoup=2]	-63,026	13,195	-91,531	-34,520	,940
[usecoup=3]	-31,375	9,726	-52,387	-10,363	,564
[usecoup=4]	,000[a]	.	.	.	.
[shopfor=1] * [usecoup=1]	41,693	11,170	17,562	65,824	,606
[shopfor=1] * [usecoup=2]	44,505	18,068	5,471	83,539	1,413
[shopfor=1] * [usecoup=3]	9,204	11,057	-14,684	33,092	,594
[shopfor=1] * [usecoup=4]	,000[a]	.	.	.	.
[shopfor=2] * [usecoup=1]	89,211	10,967	65,518	112,903	,533
[shopfor=2] * [usecoup=2]	54,267	14,949	21,972	86,562	,836
[shopfor=2] * [usecoup=3]	17,884	13,753	-11,828	47,595	,797
[shopfor=2] * [usecoup=4]	,000[a]	.	.	.	.
[shopfor=3] * [usecoup=1]	,000[a]	.	.	.	.
[shopfor=3] * [usecoup=2]	,000[a]	.	.	.	.
[shopfor=3] * [usecoup=3]	,000[a]	.	.	.	.
[shopfor=3] * [usecoup=4]	,000[a]	.	.	.	.

a. Auf 0 gesetzt, da dieser Parameter redundant ist.

b. Modell: Amount spent = (Konstanter Term) + shopfor + usecoup + shopfor * usecoup

Die Parameterschätzer zeigen den Effekt der einzelnen Einflussvariablen auf *Amount spent* (Ausgegebener Betrag). Der Wert 518.249 für den konstanten Term deutet an, dass die Einzelhandelskette davon ausgehen kann, dass ein Kunde mit Familie, der Coupons aus der Zeitung und gezielten Postsendungen verwendet, durchschnittlich $ 518,25 ausgibt. Es lässt sich sagen, dass der konstante Term diesen Faktorstufen zugeordnet ist, da es sich hierbei um die Faktorstufen handelt, deren Parameter redundant sind.

- Die *shopfor*-Koeffizienten legen nahe, dass von den Kunden, die sowohl mit der Post versandte Coupons als auch Coupons aus Zeitungen verwenden, diejenigen ohne Familie weniger ausgeben als diejenigen mit Ehepartnern, die wiederum weniger ausgeben als Personen mit unterhaltsberechtigten Familienangehörigen im Haushalt. Da die Tests der Modelleffekte zeigten, dass dieser Term zum Modell beiträgt, sollten sich diese Unterschiede nicht ändern.
- Die *usecoup*-Koeffizienten legen nahe, dass die Ausgaben bei Kunden mit unterhaltsberechtigten Familienangehörigen im Haushalt mit sinkender Coupon-Verwendung abnehmen. Es liegt ein mäßiger Grad an Unsicherheit bei den Schätzern vor, doch die Konfidenzintervalle enthalten nicht den Wert 0.
- Die Interaktionskoeffizienten legen nahe, dass Kunden, die keine Coupons oder nur Coupons aus der Zeitung verwenden und keine unterhaltsberechtigten Familienangehörigen im Haushalt haben, tendenziell mehr ausgeben, als ansonsten zu erwarten gewesen wäre. Wenn

ein Teil eines Wechselwirkungsparameters redundant ist, ist der Wechselwirkungsparameter ebenfalls redundant.

- Die Abweichung in den Werten der Effekte des Stichprobenplans vom Wert 1 weisen darauf hin, dass einige der für diese Parameterschätzer berechneten Standardfehler größer (und andere dagegen kleiner) sind als diejenigen, die man erhalten würde, wenn man davon ausginge, dass diese Beobachtungen aus einer einfachen Zufallsstichprobe stammen. Es ist von entscheidender Bedeutung, die Informationen über den Stichprobenplan in die Analyse aufzunehmen, da Sie anderenfalls beispielsweise schließen könnten, dass der Koeffizient *usecoup=3* sich nicht von 0 unterscheidet!

Die Parameterschätzer sind sinnvoll für die Quantifizierung des Effekts der einzelnen Modellterme, doch die Tabellen für die geschätzten Randmittel können die Interpretation der Modellergebnisse erleichtern.

Geschätzte Randmittel

Abbildung 19-9
Geschätzte Randmittel für die Stufen von "Who shopping for" (Einkauf für wen)

Who shopping for	Mittelwert	Standard-fehler	95%-Konfidenzintervall	
			Untere Grenze	Obere Grenze
Self	308,5326	3,94286	300,0145	317,0506
Self and spouse	370,3361	4,87908	359,7955	380,8767
Self and family	459,4392	7,19769	443,8895	474,9888

Diese Tabelle zeigt die vom Modell geschätzten Randmittel als Standardfehler von *Amount spent* (Ausgegebener Betrag) bei den Faktorstufen von *Who shopping for* (Einkauf für wen) an. Diese Tabelle kann bei der Ermittlung der Unterschiede zwischen den Stufen dieses Faktors nützlich sein. In diesem Beispiel wird erwartet, dass ein Kunde, der für sich selbst einkauft, ungefähr $ 308,53 ausgibt, während erwartet wird, dass ein Kunde mit Ehepartner $ 370,34 und ein Kunde mit unterhaltsberechtigten Familienangehörigen $ 459,44 ausgibt. Um herauszufinden, ob dies einen tatsächlichen Unterschied darstellt oder auf zufällige Schwankungen zurückzuführen ist, überprüfen Sie die Testergebnisse.

Abbildung 19-10
Einzeltestergebnisse für geschätzte Randmittel für die Geschlechter

Who shopping for Einfacher Kontrast[a]	Kontrast-schätzer	Hypothetischer Wert	Differenz (Schätzer - hypothetisch)	Standard-fehler	Freiheits-grade 1	Freiheits-grade 2	Wald-F	Sig.
Niveau Self vs. Niveau Self and family	-150,907	,000	-150,907	4,903	1,000	13,000	947,409	,000
Niveau Self and spouse vs. Niveau Self and family	-89,103	,000	-89,103	5,903	1,000	13,000	227,842	,000

a. Weggelassene Kategorie = Self and family

Die Tabelle mit den Einzeltests zeigt zwei einfache Kontraste bei den ausgegebenen Beträgen.

- Der Kontrastschätzer ist die Differenz bei den ausgegebenen Beträgen für die aufgeführten Stufen von *Who shopping for* (Einkauf für wen).

- Der hypothetische Wert 0,00 steht für die Vermutung, dass keine Unterschiede bei den ausgegebenen Beträgen vorliegen.
- Die Wald *F*-Statistik mit den angezeigten Freiheitsgraden wird verwendet, um zu testen, ob die Differenz zwischen einem Kontrastschätzer und einem hypothetischen Wert auf zufällige Schwankungen zurückzuführen ist.
- Da die Signifikanzwerte unter 0,05 liegen, können Sie davon ausgehen, dass Unterschiede in den ausgegebenen Beträgen vorliegen.

Die Werte der Kontrastschätzer weichen von den Parameterschätzern ab. Dies liegt daran, dass ein Wechselwirkungs-Term vorliegt, der den Effekt *Who shopping for* (Einkauf für wen) enthält. Als Ergebnis ist der Parameterschätzer für *shopfor=1* ein einfacher Kontrast zwischen den Stufen *Self* (Eigene Person) und *Self and Family* (Eigene Person und Familie) auf der Stufe *From both* (Aus beiden Quellen) der Variablen *Use coupons* (Verwendung von Coupons). Für den Kontrastschätzer in dieser Tabelle wird der Durchschnitt über die einzelnen Stufen von *Use coupons* (Verwendung von Coupons) ermittelt.

Abbildung 19-11
Ergebnisse des Gesamttests für geschätzte Randmittel für die Geschlechter

Freiheits-grade 1	Freiheits-grade 2	Wald-F	Sig.
2,000	12,000	643,593	,000

In der Tabelle mit dem Gesamttest finden Sie die Ergebnisse eines Tests aller Kontraste in der Einzeltesttabelle. Der Signifikanzwert von weniger als “0,05” bestätigt, dass ein Unterschied in den ausgegebenen Beträgen zwischen den einzelnen Stufen von *Who shopping for* (Einkauf für wen) vorliegt.

Abbildung 19-12
Geschätzte Randmittel für die Stufen des Einkaufsstils

			95%-Konfidenzintervall	
Use coupons	Mittelwert	Standard-fehler	Untere Grenze	Obere Grenze
No	319,6455	6,51429	305,5722	333,7188
From newspaper	386,7469	4,32295	377,4077	396,0861
From mailings	394,5028	5,54218	382,5297	406,4760
From both	416,8486	6,51260	402,7790	430,9182

Diese Tabelle zeigt die vom Modell geschätzten Randmittel und Standardfehler von *Amount spent* (Ausgegebener Betrag) bei den Faktorstufen von *Use coupons* (Verwendung von Coupons) an. Diese Tabelle kann bei der Ermittlung der Unterschiede zwischen den Stufen dieses Faktors nützlich sein. In diesem Beispiel wird erwartet, dass ein Kunde ungefähr $ 319.65 ausgibt, und es wird erwartet, dass diejenigen, die Coupons verwenden, erheblich mehr ausgeben.

Abbildung 19-13
Einzeltestergebnisse für geschätzte Randmittel des Einkaufsstils

Use coupons Einfacher Kontrast[a]	Kontrast-schätzer	Hypothetischer Wert	Differenz (Schätzer - hypothetisch)	Standard-fehler	Freiheits-grade 1	Freiheits-grade 2	Wald-F	Sig.
Niveau From newspaper vs. Niveau No	67,101	,000	67,101	6,537	1,000	13,000	105,352	,000
Niveau From mailings vs. Niveau No	74,857	,000	74,857	5,875	1,000	13,000	162,328	,000
Niveau From both vs. Niveau No	97,203	,000	97,203	5,603	1,000	13,000	300,921	,000

a. Weggelassene Kategorie = No

Die Tabelle mit den Einzeltests zeigt drei einfache Kontraste an, in der das Ausgabeverhalten der Kunden, die keine Coupons verwenden, mit dem der Kunden verglichen wird, die Coupons einsetzen.

Da die Signifikanzwerte der Tests unter 0,05 liegen, können Sie schließen, dass Kunden, die Coupons verwenden, in der Regel mehr ausgeben als diejenigen, die keine Coupons verwenden.

Abbildung 19-14
Ergebnisse des Gesamttests für geschätzte Randmittel des Einkaufsstils

Freiheits-grade 1	Freiheits-grade 2	Wald-F	Sig.
3,000	11,000	87,453	,000

In der Tabelle mit dem Gesamttest finden Sie die Ergebnisse eines Tests der Kontraste in der Einzeltesttabelle. Der Signifikanzwert von weniger als “0,05” bestätigt, dass ein Unterschied in den ausgegebenen Beträgen zwischen den einzelnen Stufen von *Use coupons* (Verwendung von Coupons) vorliegt. Beachten Sie: Die Gesamttests für *Use coupons* (Verwendung von Coupons) und *Who shopping for* (Einkauf für wen) sind mit den Tests der Modelleffekte äquivalent, da die hypothetischen Kontrastwerte gleich 0 sind.

Abbildung 19-15
Geschätzte Randmittel für die Stufen von Geschlecht nach Einkaufsstil

				95%-Konfidenzintervall	
Who shopping for	Use coupons	Mittelwert	Standard-fehler	Untere Grenze	Obere Grenze
Self	No	244,3471	6,00949	231,3644	257,3298
	From newspaper	324,9708	5,94134	312,1353	337,8063
	From mailings	321,3207	4,11028	312,4410	330,2005
	From both	343,4916	6,57845	329,2797	357,7034
Self and spouse	No	337,1783	7,12181	321,7925	352,5640
	From newspaper	380,0468	7,91038	362,9574	397,1361
	From mailings	375,3141	6,22468	361,8665	388,7617
	From both	388,8054	7,12101	373,4214	404,1894
Self and family	No	377,4111	11,58215	352,3894	402,4328
	From newspaper	455,2232	6,14420	441,9494	468,4969
	From mailings	486,8736	10,76529	463,6166	510,1306
	From both	518,2488	11,73120	492,9050	543,5925

Diese Tabelle zeigt die vom Modell geschätzten Randmittel, Standardfehler und Konfidenzintervalle von *Amount spent* (Ausgegebener Betrag) bei den Faktorkombinationen von *Who shopping for* (Einkauf für wen) und *Use coupons* (Coupons verwenden) an. Diese Tabelle

dient zur Ermittlung des Wechselwirkungseffekts zwischen diesen beiden Faktoren, der in den Tests der Modelleffekte gefunden wurde.

Zusammenfassung

In diesem Beispiel ergaben die geschätzten Randmittel Unterschiede in den ausgegebenen Beträgen zwischen Kunden auf verschiedenen Stufen von *Who shopping for* (Einkauf für wen) und *Use coupons* (Coupons verwenden). Dies wurde durch die Tests der Modelleffekte bestätigt sowie durch die Tatsache, dass ein Wechselwirkungseffekt *Who shopping for*Use coupons* vorzuliegen scheint. Aus der Modellzusammenfassungstabelle ergab sich, dass das vorliegende Modell etwas mehr als die Hälfte der Schwankungen in den Daten erklärt und vermutlich durch die Hinzunahme weiterer Einflussvariablen verbessert werden könnte.

Verwandte Prozeduren

Die Prozedur “Allgemeines lineares Modell für komplexe Stichproben” ist ein nützliches Tool für die Modellierung einer metrischen Variablen, wenn die Fälle anhand eines Schemas für komplexe Stichproben gezogen wurden.

- Der Stichprobenassistent für komplexe Stichproben wird zur Angabe der Planspezifikationen für komplexe Stichproben und zum Ziehen von Stichproben verwendet. Die vom Stichprobenassistenten erstellte Stichprobenplan-Datei enthält einen Standard-Analyseplan und kann im Dialogfeld “Plan” angegeben werden, wenn die gezogene Stichprobe gemäß diesem Plan analysiert werden soll.
- Der Analysevorbereitungsassistent für komplexe Stichproben wird zur Angabe der Analysespezifikationen für eine bestehende komplexe Stichprobe verwendet. Die vom Stichprobenassistenten erstellte Analyseplan-Datei kann im Dialogfeld “Plan” angegeben werden, wenn Sie die Stichprobe gemäß diesem Plan analysieren.
- Die Prozedur Logistische Regression für komplexe Stichproben ermöglicht die Modellierung einer kategorialen Antwort (Responsevariablen).
- Die Prozedur Ordinale Regression für komplexe Stichproben ermöglicht die Modellierung einer ordinalen Antwort (Responsevariablen).

Kapitel
20

Logistische Regression für komplexe Stichproben

Die Prozedur “Logistische Regression für komplexe Stichproben” besteht aus einer logistischen Regressionsanalyse einer binären oder multinomialen abhängigen Variable für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Optional können Sie auch Analysen für eine Teilgesamtheit vornehmen.

Verwenden der logistischen Regression für komplexe Stichproben zur Bewertung des Kreditrisikos

Als Kreditsachbearbeiter in einer Bank sollten Sie in der Lage sein, Merkmale zu ermitteln, die auf Personen hindeuten, die mit hoher Wahrscheinlichkeit ihre Kredite nicht zurückzahlen, und diese Merkmale zur Feststellung eines guten bzw. schlechten Kreditrisikos einzusetzen.

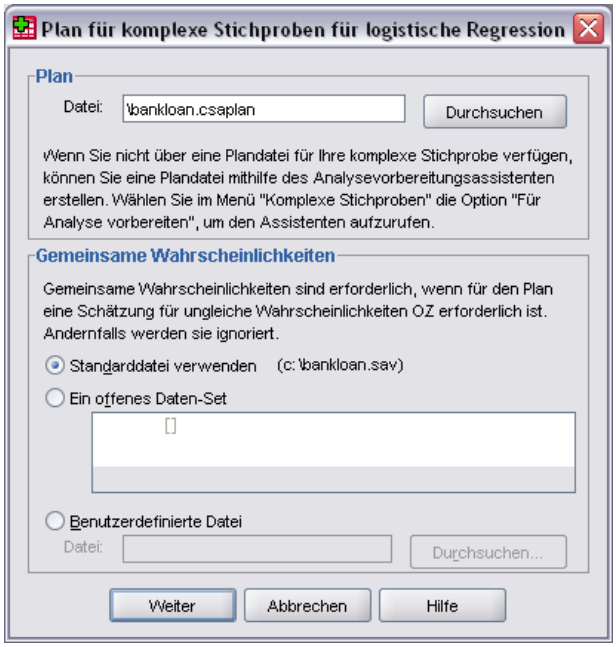
Angenommen, ein Kreditsachbearbeiter verfügt über eine Reihe von Datensätzen zu Kunden, die ein Darlehen in verschiedenen Zweigstellen erhalten haben; diese Datensätze wurden nach einem komplexen Plan zusammengestellt. Diese Informationen finden Sie in *bankloan_cs.sav*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Der Sachbearbeiter interessiert sich für die Wahrscheinlichkeit, mit der ein Kunde einen Kredit nicht zurückzahlen kann, und zwar im Zusammenhang mit dem Alter, der beruflichen Entwicklung und der Höhe des Kredits. Bei dieser Untersuchung soll der Stichprobenplan berücksichtigt werden.

Durchführung der Analyse

▶ Wählen Sie zum Erstellen des logistischen Regressionsmodells die folgenden Menübefehle aus:

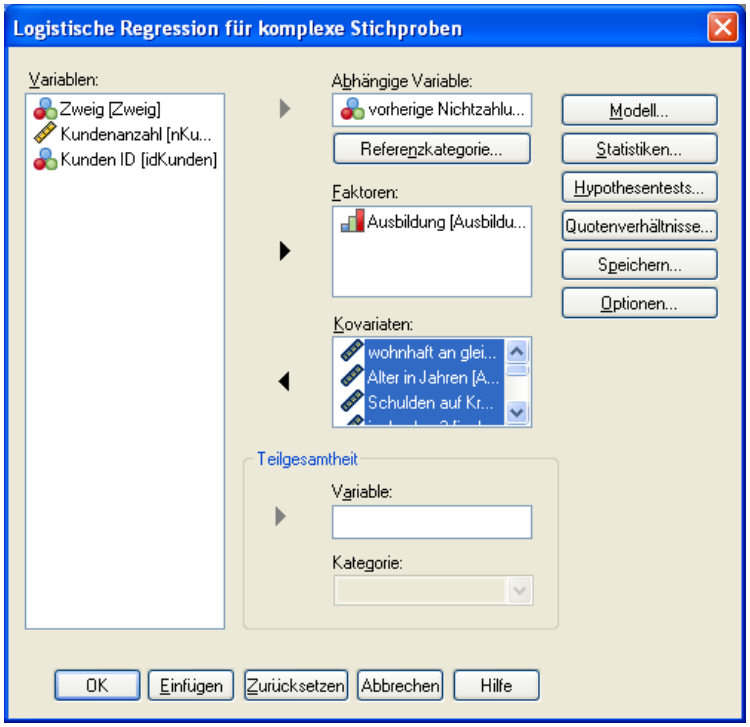
Analysieren
  Komplexe Stichproben
    Logistische Regression...

Abbildung 20-1
Dialogfeld "Komplexe Stichproben: Plan"


Plan für komplexe Stichproben für logistische Regression
Plan
Datei:
\bankloan.csaplan
Durchsuchen
Wenn Sie nicht über eine Plandatei für Ihre komplexe Stichprobe verfügen, können Sie eine Plandatei mithilfe des Analysevorbereitungsassistenten erstellen. Wählen Sie im Menü "Komplexe Stichproben" die Option "Für Analyse vorbereiten", um den Assistenten aufzurufen.
Gemeinsame Wahrscheinlichkeiten
Gemeinsame Wahrscheinlichkeiten sind erforderlich, wenn für den Plan eine Schätzung für ungleiche Wahrscheinlichkeiten OZ erforderlich ist. Andernfalls werden sie ignoriert.
Standarddatei verwenden (c:\bankloan.sav)
Ein offenes Daten-Set
Benutzerdefinierte Datei
Datei:
Durchsuchen...
Weiter
Abbrechen
Hilfe

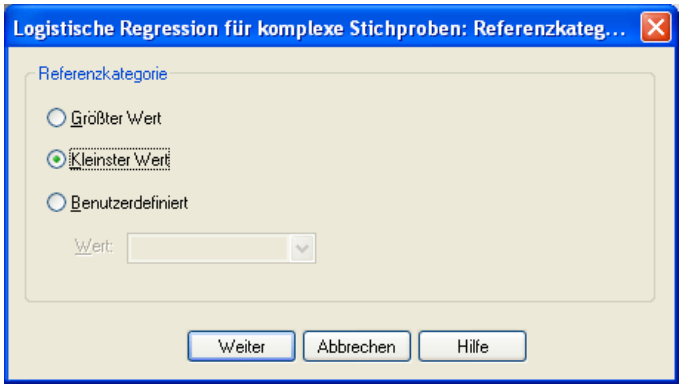
- Wechseln Sie zu der Datei *bankloan.csaplan* und wählen Sie sie aus. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271.
- Klicken Sie auf Weiter.

Abbildung 20-2
Dialogfeld "Logistische Regression"



- Wählen Sie *Vorherige Nichtzahlung* als abhängige Variable aus.
- Wählen Sie *Ausbildung* als Faktor aus.
- Wählen Sie *Alter in Jahren* bis *Andere Schulden in Tausend* als Kovariaten aus.
- Wählen Sie *Vorherige Nichtzahlung* aus und klicken Sie auf Referenzkategorie.

Abbildung 20-3
Dialogfeld "Logistische Regression: Referenzkategorie"



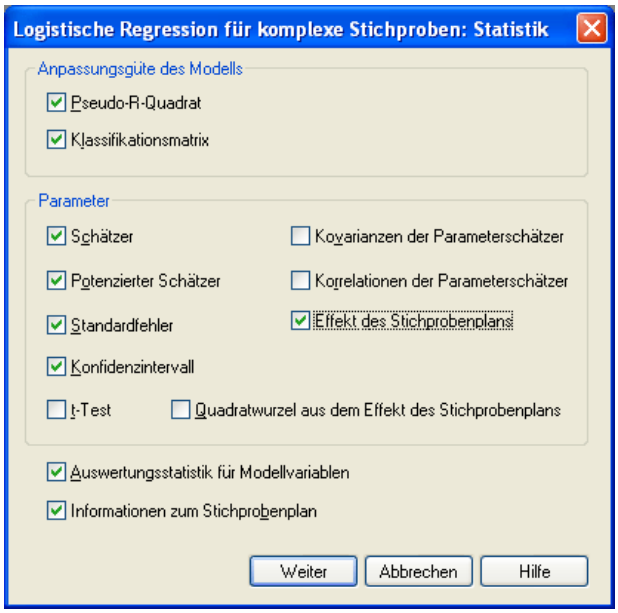
- Wählen Sie Kleinster Wert als Referenzkategorie aus.

  Dadurch wird die Kategorie "keine Nichtzahlung" als Referenzkategorie festgelegt. Die in der Ausgabe enthaltenen Quotenverhältnisse haben daher die Eigenschaft, dass steigende Quotenverhältnisse einer höheren Wahrscheinlichkeit einer Nichtzahlung entsprechen.

- Klicken Sie auf Weiter.

- Klicken Sie im Dialogfeld "Logistische Regression" auf Statistik.

Abbildung 20-4
Dialogfeld "Logistische Regression: Statistik"



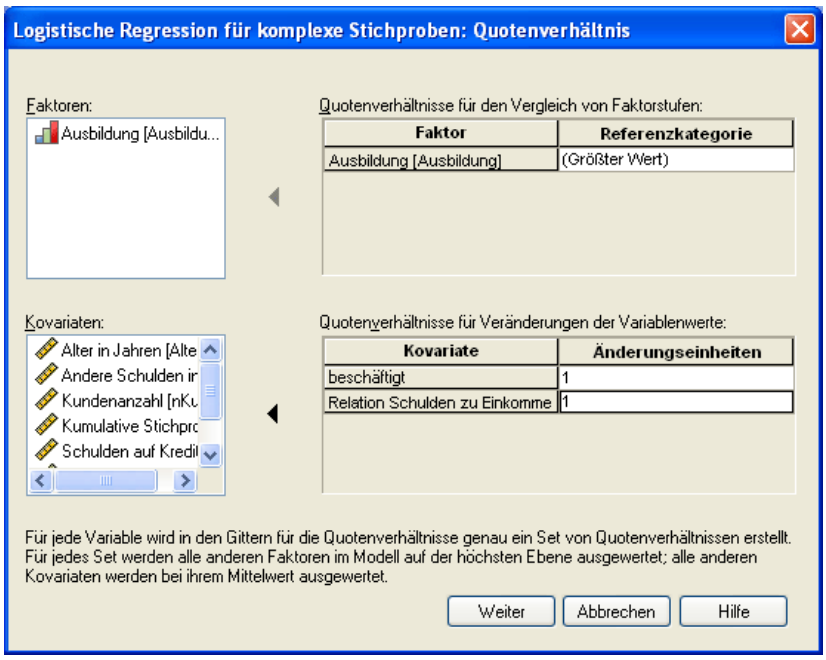
- Wählen Sie im Gruppenfeld "Anpassungsgüte des Modells" die Option Klassifikationsmatrix.

- Wählen Sie in der Gruppe "Parameter" die Optionen Schätzer, Exponentialfunktion des Schätzers, Standardfehler, Konfidenzintervall und Effekt des Stichprobenplans aus.

- Klicken Sie auf Weiter.

- Klicken Sie im Dialogfeld "Logistische Regression" auf Quotenverhältnisse.

Abbildung 20-5
Dialogfeld "Logistische Regression: Quotenverhältnisse"



- Legen Sie fest, dass Quotenverhältnisse für den Faktor *Ausbildung* und die Kovariaten *beschäftigt* und *Schulden_Eink* erstellt werden sollen.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Logistische Regression" auf OK.

Werte für Pseudo-R-Quadrat

Abbildung 20-6
Pseudo-R-Quadrat-Statistik

Cox und Snell	,330
Nagelkerke	,451
McFadden	,304

Abhängige Variable: vorherige Nichtzahlung (Referenzkategorie = Nein)
Modell: (Konstanter Term), Ausbildung, Alter, beschäftigt, wohnhaft, Einkommen, Schulden_Eink, Card_Schulden, sonst_Schulden

Im linearen Regressionsmodell fasst das Bestimmtheitsmaß, R^2, den Anteil der Varianz in der abhängigen Variable, das mit den (unabhängigen) Einflussvariablen zu tun hat, zusammen. Dabei deuten größere R^2-Werte darauf hin, dass ein größerer Anteil der Varianz durch das Modell erklärt wird, bis zu einem Maximalwert von 1. Bei Regressionsmodellen mit einer kategorialen abhängigen Variablen kann keine einzelne R^2-Statistik berechnet werden, die alle Merkmale

von R^2 im linearen Regressionsmodell aufweist. Daher werden stattdessen diese Näherungen berechnet. Folgende Verfahren werden verwendet, um das Bestimmtheitsmaß abzuschätzen.

- R^2 nach Cox und Snell (Cox als auch Snell, 1989) beruht auf der Log-Likelihood für das Modell im Vergleich mit der Log-Likelihood für ein Grundlinienmodell. Bei kategorialen Ergebnissen hat es jedoch einen theoretischen Maximalwert von weniger als 1, sogar für ein "perfektes" Modell.
- R^2 nach Nagelkerke (Nagelkerke, 1991) ist eine korrigierte Version des *R*-Quadrats nach Cox & Snell, bei dem die Skala der Statistik so angepasst wird, dass sie den vollständigen Bereich von 0 bis 1 abdeckt.
- R^2 nach McFadden (McFadden, 1974) ist eine weitere Version, die auf den Log-Likelihood-Kernels für das Modell mit ausschließlich konstanten Termen und das vollständige geschätzte Modell beruht.

Was einen "guten" R^2-Wert ausmacht, hängt von den verschiedenen Anwendungsbereichen ab. Diese Statistiken können zwar auch für sich genommen bereits Schlüsse erlauben, sie sind jedoch am sinnvollsten, wenn es um den Vergleich von konkurrierenden Modellen für dieselben Daten geht. Das Modell mit der größten R^2-Statistik ist nach diesem Maßstab am "besten".

Klassifikation

Abbildung 20-7
Klassifikationsmatrix

	Vorhergesagt		
Beobachtet	Nein	ja	Prozent korrekt
Nein	188289,7	31871,267	85,5%
ja	49970,600	77675,133	60,9%
Prozent (insgesamt)	68,5%	31,5%	76,5%

Abhängige Variable: vorherige Nichtzahlung
(Referenzkategorie = Nein)
Modell: (Konstanter Term), Ausbildung, Alter, beschäftigt, wohnhaft, Einkommen, Schulden_Eink, Card_Schulden, sonst_Schulden

Die Klassifikationsmatrix zeigt die praktischen Ergebnisse der Verwendung des logistischen Regressionsmodells. In jedem Fall ist die vorhergesagte Antwort *Ja*, wenn der vom Modell vorhergesagte Logit-Wert größer als 0 ist. Die Fälle werden nach *finalweight* gewichtet, sodass die Klassifikationsmatrix die erwartete Modellleistung in der Grundgesamtheit wiedergibt.

- Die Zellen auf der Diagonale stellen korrekte Vorhersagen dar.
- Die Zellen abseits der Diagonale stellen falsche Vorhersagen dar.

Auf der Grundlage der zum Erstellen des Modells verwendeten Fälle können Sie davon ausgehen, dass Sie mit diesem Modell 85,5 % der Personen in der Grundgesamtheit, die nicht zahlungsunfähig werden, korrekt klassifizieren. Ebenso können Sie davon ausgehen, dass Sie 60,9 % der Personen, die zahlungsunfähig werden, korrekt klassifizieren. Insgesamt können Sie davon ausgehen, dass 76,5 % der Fälle korrekt klassifiziert werden. Da diese Tabelle jedoch mit den Fällen erstellt wurde, die auch zum Aufbau des Modells dienten, sind diese Schätzungen wahrscheinlich zu optimistisch.

Tests der Modelleffekte

Abbildung 20-8
Tests der Zwischensubjekteffekte

Quelle	Freiheits-grade 1	Freiheits-grade 2	Wald-F	Sig.
(Korrigiertes Modell)	11,000	4,000	14,669	,010
(Konstanter Term)	1,000	14,000	5,777	,031
Ausbildung	4,000	11,000	1,683	,224
Alter	1,000	14,000	5,352	,036
beschäftigt	1,000	14,000	88,244	,000
wohnhaft	1,000	14,000	1,123	,307
Einkommen	1,000	14,000	,007	,932
Schulden_Eink	1,000	14,000	27,632	,000
Card_Schulden	1,000	14,000	33,402	,000
sonst_Schulden	1,000	14,000	,709	,414

Abhängige Variable: vorherige Nichtzahlung (Referenzkategorie = Nein)
Modell: (Konstanter Term), Ausbildung, Alter, beschäftigt, wohnhaft, Einkommen, Schulden_Eink, Card_Schulden, sonst_Schulden

Jeder Term im Modell sowie das Modell als Ganzes werden daraufhin getestet, ob sein Effekt gleich 0 ist. Terme mit Signifikanzwerten von weniger als 0,05 weisen einen erkennbaren Effekt auf. So tragen *Alter*, *beschäftigt*, *Schulden_Eink* und *Card_Schulden* zum Modell bei, wohingegen die andern Haupteffekte keinen Beitrag zum Modell leisten. Bei einer weiteren Analyse der Daten würden Sie vermutlich *Ausbildung*, *wohnhaft*, *Einkommen* und *sonst_Schulden* aus der Modellbetrachtung herausnehmen.

Parameterschätzer

Abbildung 20-9
Parameterschätzer

vorherige Nichtzahlung	Parameter	B	Standard-fehler	95%-Konfidenzintervall		Exp(B)	95% Konfidenzintervall für Exp(B)	
				Untere Grenze	Obere Grenze		Untere Grenze	Obere Grenze
ja	(Konstanter Term)	-1,140	,399	-1,995	-,284	,320	,136	,753
	[Ausbildung=1]	,720	,340	-,010	1,449	2,054	,990	4,259
	[Ausbildung=2]	,684	,371	-,112	1,481	1,983	,894	4,397
	[Ausbildung=3]	,518	,307	-,140	1,177	1,679	,869	3,244
	[Ausbildung=4]	,789	,302	,142	1,437	2,202	1,152	4,208
	[Ausbildung=5]	,000[a]	.	.	.	1,000	.	.
	Alter	-,023	,010	-,043	-,002	,978	,958	,998
	beschäftigt	-,225	,024	-,277	-,174	,798	,758	,840
	wohnhaft	-,028	,026	-,085	,029	,972	,919	1,029
	Einkommen	,000	,003	-,007	,006	1,000	,993	1,006
	Schulden_Eink	,095	,018	,056	,134	1,100	1,058	1,143
	Card_Schulden	,493	,085	,310	,676	1,637	1,363	1,966
	sonst_Schulden	,026	,031	-,041	,094	1,027	,960	1,098

Abhängige Variable: vorherige Nichtzahlung (Referenzkategorie = Nein)
Modell: (Konstanter Term), Ausbildung, Alter, beschäftigt, wohnhaft, Einkommen, Schulden_Eink, Card_Schulden, sonst_Schulden

a. Auf 0 gesetzt, da dieser Parameter redundant ist.

Die Tabelle der Parameterschätzer fasst den Effekt der einzelnen Einflussvariablen zusammen. Beachten Sie: Die Parameterwerte beeinflussen die relative Wahrscheinlichkeit der Kategorie "Nichtzahlung" gegenüber der Kategorie "keine Nichtzahlung". So erhöhen Parameter mit

positiven Koeffizienten die Wahrscheinlichkeit einer Nichtzahlung, wohingegen Parameter mit negativen Koeffizienten die Wahrscheinlichkeit der Nichtzahlung verringern.

Die Bedeutung von logistischen Regressionskoeffizienten ist nicht so eindeutig wie die von linearen Regressionskoeffizienten. Während *B* praktisch zum Testen der Modelleffekte ist, ist *Exp(B)* einfacher zu interpretieren. *Exp(B)* steht für die Änderung in den Quotenverhältnissen des zu untersuchenden Ereignisses, die auf einen Anstieg um eine Einheit für Einflussvariablen zurückzuführen ist, die nicht Teil von Wechselwirkungstermen sind. Beispiel: *Exp(B)* für *beschäftigt* entspricht 0,798, was bedeutet, dass die Quote für Nichtzahlung bei Personen, die seit zwei Jahren bei ihrem derzeitigen Arbeitgeber beschäftigt sind, das 0,798fache der Quote für Nichtzahlung bei den Personen beträgt, die seit einem Jahr bei ihrem derzeitigen Arbeitgeber beschäftigt sind, sofern alle anderen Faktoren gleich sind.

Die Effekte des Stichprobenplans weisen darauf hin, dass einige der für diese Parameterschätzer berechneten Standardfehler größer (und andere dagegen kleiner) sind als diejenigen, die man erhalten würde, wenn man davon ausginge, dass diese Beobachtungen aus einer einfachen Zufallsstichprobe stammen. Es ist von entscheidender Bedeutung, die Informationen über den Stichprobenplan in die Analyse aufzunehmen, da Sie anderenfalls beispielsweise schließen könnten, dass der Koeffizient “Alter” sich nicht von 0 unterscheidet.

Quotenverhältnisse

Abbildung 20-10
Quotenverhältnisse für “Ausbildung”

		vorherige Nichtzahlung	Quoten-verhältnis	95%-Konfidenzintervall	
				Untere Grenze	Obere Grenze
Ausbildung	ohne Schulabschluß vs. Promotion	ja	2,054	,990	4,259
	Hauptschul-/	ja	1,983	,894	4,397
	Some college vs.	ja	1,679	,869	3,244
	Universitätsabschluß vs.	ja	2,202	1,152	4,208

Abhängige Variable: vorherige Nichtzahlung (Referenzkategorie = Nein)
Modell: (Konstanter Term), Ausbildung, Alter, beschäftigt, wohnhaft, Einkommen, Schulden_Eink, Card_ Schulden, sonst_Schulden

a. Die bei der Berechnung verwendeten Faktoren und Kovariaten werden auf folgende Werte festgesetzt: Ausbildung=Promotion; Alter in Jahren=34,19; Jahre der Beschäftigung beim derzeitigen Arbeitgeber=6,99; wohnhaft an gleicher Adresse (in Jahren)=6,32; Haushaltseinkommen in Tausend=60,1581; Relation Schulden zu Einkommen (in %)=9,9341; Schulden auf Kreditkarte in Tausend=1,9764; Andere Schulden in Tausend=3,9164

Diese Tabelle enthält die Quotenverhältnisse von *Vorherige Nichtzahlung* auf den Faktorstufen von *Ausbildung*. Bei den ausgegebenen Werten handelt es sich um das Verhältnis der Quoten für die Nichtzahlung für *Ohne Schulabschluss* bis *Universitätsabschluss* im Vergleich zu der Quote für die Nichtzahlung für *Promotion*. Das Quotenverhältnis von 2,054 in der ersten Zeile der Tabelle bedeutet, dass die Quote für die Nichtzahlung bei einer Person ohne Schulabschluss das 2,054fache der Quote für die Nichtzahlung bei einer promovierten Person betägt.

Abbildung 20-11
Quotenverhältnisse für "Jahre der Beschäftigung beim derzeitigen Arbeitgeber"

Änderungseinheiten		vorherige Nichtzahlung	Quoten-verhältnis	95%-Konfidenzintervall	
				Untere Grenze	Obere Grenze
Jahre der Beschäftigung beim derzeitigen Arbeitgeber	1,000	ja	,798	,758	,840

Abhängige Variable: vorherige Nichtzahlung (Referenzkategorie = Nein)
Modell: (Konstanter Term), Ausbildung, Alter, beschäftigt, wohnhaft, Einkommen, Schulden_Eink, Card_Schulden, sonst_Schulden

a. Die bei der Berechnung verwendeten Faktoren und Kovariaten werden auf folgende Werte festgesetzt: Ausbildung=Promotion; Alter in Jahren=34,19; Jahre der Beschäftigung beim derzeitigen Arbeitgeber=6,99; wohnhaft an gleicher Adresse (in Jahren)=6,32; Haushaltseinkommen in Tausend=60,1581; Relation Schulden zu Einkommen (in %)=9,9341; Schulden auf Kreditkarte in Tausend=1,9764; Andere Schulden in Tausend=3,9164

Diese Tabelle enthält das Quotenverhältnis von *Vorherige Nichtzahlung* für eine Einheitsänderung in der Kovariate *Jahre der Beschäftigung beim derzeitigen Arbeitgeber*. Bei dem ausgegebenen Wert handelt es sich um das Verhältnis aus der Quote für die Nichtzahlung für eine Person, die bereits seit 7,99 Jahren an ihrem derzeitigen Arbeitsplatz beschäftigt ist, im Vergleich zu der Quote für die Nichtzahlung für eine Person mit 6,99 Jahren (Mittelwert).

Abbildung 20-12
Quotenverhältnisse für "Relation Schulden zu Einkommen"

Änderungseinheiten		vorherige Nichtzahlung	Quoten-verhältnis	95%-Konfidenzintervall	
				Untere Grenze	Obere Grenze
Relation Schulden zu Einkommen (in %)	1,000	ja	1,100	1,058	1,143

Abhängige Variable: vorherige Nichtzahlung (Referenzkategorie = Nein)
Modell: (Konstanter Term), Ausbildung, Alter, beschäftigt, wohnhaft, Einkommen, Schulden_Eink, Card_Schulden, sonst_Schulden

a. Die bei der Berechnung verwendeten Faktoren und Kovariaten werden auf folgende Werte festgesetzt: Ausbildung=Promotion; Alter in Jahren=34,19; Jahre der Beschäftigung beim derzeitigen Arbeitgeber=6,99; wohnhaft an gleicher Adresse (in Jahren)=6,32; Haushaltseinkommen in Tausend=60,1581; Relation Schulden zu Einkommen (in %)=9,9341; Schulden auf Kreditkarte in Tausend=1,9764; Andere Schulden in Tausend=3,9164

Diese Tabelle enthält das Quotenverhältnis von *Vorherige Nichtzahlung* für eine Einheitsänderung in der Kovariate *Relation Schulden zu Einkommen*. Bei dem ausgegebenen Wert handelt es sich um das Verhältnis aus der Quote für die Nichtzahlung für eine Person mit einer Relation von Schulden zu Einkommen von 10,9341 im Vergleich zu der Quote für die Nichtzahlung für eine Person mit einem Wert von 9,9341 (Mittelwert).

Beachten Sie: Da keine dieser Einflussvariablen Teil von Wechselwirkungstermen sind, sind die Werte der in diesen Tabellen enthaltenen Quotenverhältnisse gleich der Werte der potenzierten Parameterschätzer. Wenn eine Einflussvariable Teil eines Wechselwirkungsterms ist, hängt das in diesen Tabellen verzeichnete Quotenverhältnis auch von den Werten der anderen Einflussvariablen ab, die für die Wechselwirkung verwendet werden.

Zusammenfassung

Mit der Prozedur “Logistische Regression für komplexe Stichproben” haben Sie ein Modell für die Vorhersage der Wahrscheinlichkeit erstellt, mit der ein bestimmter Kunde einen Kredit nicht zurückzahlen wird.

Eine wichtige Frage für Kreditsachbearbeiter sind die Kosten für Fehler erster und zweiter Art. Wie hoch sind die Kosten der Einstufung einer zahlungsunfähigen Person in die Gruppe der nicht zahlungsunfähigen Personen (Fehler erster Art)? Wie hoch sind die Kosten der Einstufung einer nicht zahlungsunfähigen Person in die Gruppe der zahlungsunfähigen Personen (Fehler zweiter Art)? Wenn uneinbringliche Forderungen der wichtigste Punkt sind, sollte der Fehler erster Art minimiert und die **Sensitivität** maximiert werden. Wenn die Erweiterung des Kundenstamms oberste Priorität hat, sollte der Fehler zweiter Art minimiert und die **Spezifität** maximiert werden. Normalerweise sind beide Punkte von großer Bedeutung, sodass Sie eine Entscheidungsregel für die Klassifizierung von Kunden aufstellen müssen, die die beste Mischung aus Sensitivität und Spezifität bietet.

Verwandte Prozeduren

Die Prozedur “Logistische Regression für komplexe Stichproben” ist ein nützliches Tool für die Modellierung einer kategorialen Variablen, wenn die Fälle anhand eines Schemas für komplexe Stichproben gezogen wurden.

- Der Stichprobenassistent für komplexe Stichproben wird zur Angabe der Planspezifikationen für komplexe Stichproben und zum Ziehen von Stichproben verwendet. Die vom Stichprobenassistenten erstellte Stichprobenplan-Datei enthält einen Standard-Analyseplan und kann im Dialogfeld “Plan” angegeben werden, wenn die gezogene Stichprobe gemäß diesem Plan analysiert werden soll.
- Der Analysevorbereitungsassistent für komplexe Stichproben wird zur Angabe der Analysespezifikationen für eine bestehende komplexe Stichprobe verwendet. Die vom Stichprobenassistenten erstellte Analyseplan-Datei kann im Dialogfeld “Plan” angegeben werden, wenn Sie die Stichprobe gemäß diesem Plan analysieren.
- Die Prozedur Allgemeines lineares Modell für komplexe Stichproben ermöglicht die Modellierung einer metrischen Antwort (Responsevariablen).
- Die Prozedur Ordinale Regression für komplexe Stichproben ermöglicht die Modellierung einer ordinalen Antwort (Responsevariablen).

Kapitel
21

Ordinale Regression für komplexe Stichproben

Die Prozedur "Ordinale Regression für komplexe Stichproben" erstellt ein Vorhersagemodell für eine ordinale abhängige Variable für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Optional können Sie auch Analysen für eine Teilgesamtheit vornehmen.

Verwendung der ordinalen Regression für komplexe Stichproben zur Analyse von Umfrageergebnissen

Abgeordnete, die in Erwägung ziehen, einen Gesetzesentwurf einzubringen, sind daran interessiert zu ermitteln, ob dieser Gesetzesantrag öffentlich unterstützt wird und in welchem Bezug die Unterstützung für den Antrag zur demografischen Struktur der Wähler steht. Die Meinungsforscher verwenden für die Erstellung und Durchführung der entsprechenden Umfragen einen komplexen Stichprobenplan.

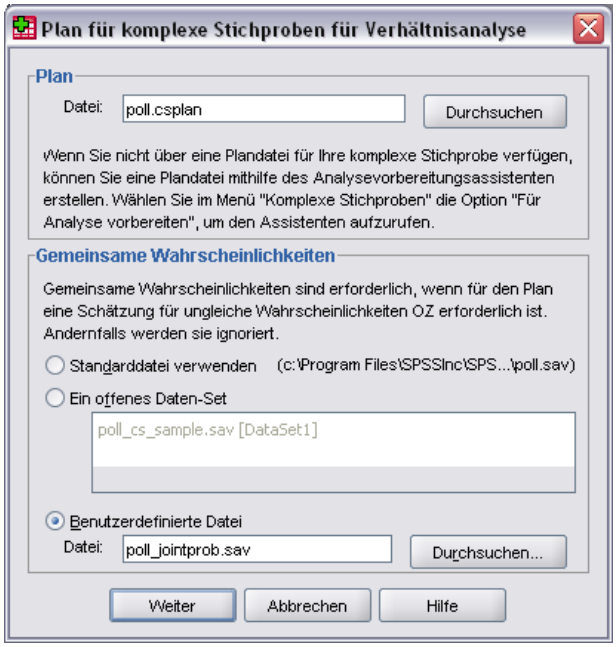
Die Umfrageergebnisse finden Sie in der Datei *poll_cs_sample.sav*. Der von den Meinungsforschern verwendete Stichprobenplan befindet sich in *poll.csplan*. Da hier die PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*poll_jointprob.sav*). Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Mit der ordinalen Regression für komplexe Stichproben können Sie ein Modell für die Stärke der Unterstützung für den Gesetzesentwurf auf der Grundlage der demografischen Struktur der Wähler anpassen.

Durchführung der Analyse

▶ Um eine Analyse der Art "Ordinale Regression für komplexe Stichproben" durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

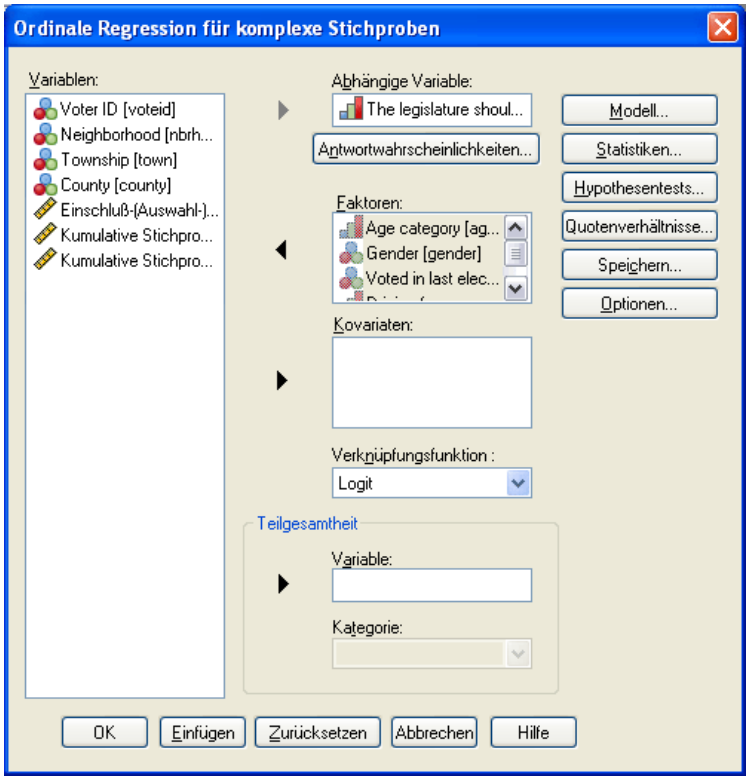
Analysieren
  Komplexe Stichproben
    Ordinale Regression...

Abbildung 21-1
Dialogfeld "Komplexe Stichproben: Plan"


Plan für komplexe Stichproben für Verhältnisanalyse
Plan
Datei:
poll.csplan
Durchsuchen
Wenn Sie nicht über eine Plandatei für Ihre komplexe Stichprobe verfügen, können Sie eine Plandatei mithilfe des Analysevorbereitungsassistenten erstellen. Wählen Sie im Menü "Komplexe Stichproben" die Option "Für Analyse vorbereiten", um den Assistenten aufzurufen.
Gemeinsame Wahrscheinlichkeiten
Gemeinsame Wahrscheinlichkeiten sind erforderlich, wenn für den Plan eine Schätzung für ungleiche Wahrscheinlichkeiten OZ erforderlich ist. Andernfalls werden sie ignoriert.
Standarddatei verwenden (c:\Program Files\SPSSInc\SPS...\poll.sav)
Ein offenes Daten-Set
poll_cs_sample.sav [DataSet1]
Benutzerdefinierte Datei
Datei:
poll_jointprob.sav
Durchsuchen...
Weiter
Abbrechen
Hilfe

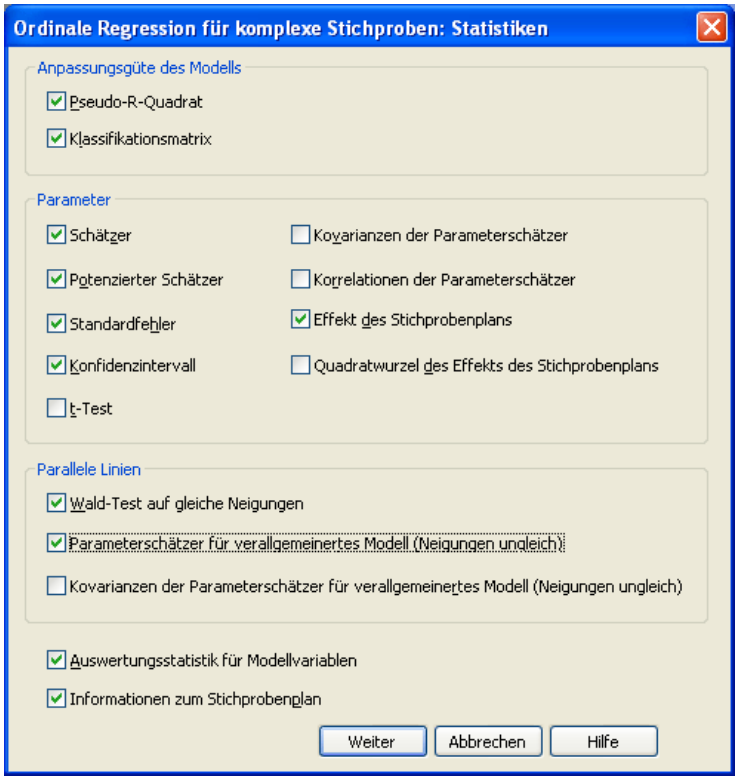
- Wechseln Sie zu der Datei *poll.csplan* und wählen Sie sie als Plandatei aus. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271.
- Wählen Sie *poll_jointprob.sav* als Datei für die gemeinsamen Wahrscheinlichkeiten.
- Klicken Sie auf Weiter.

Abbildung 21-2
Dialogfeld "Ordinale Regression"



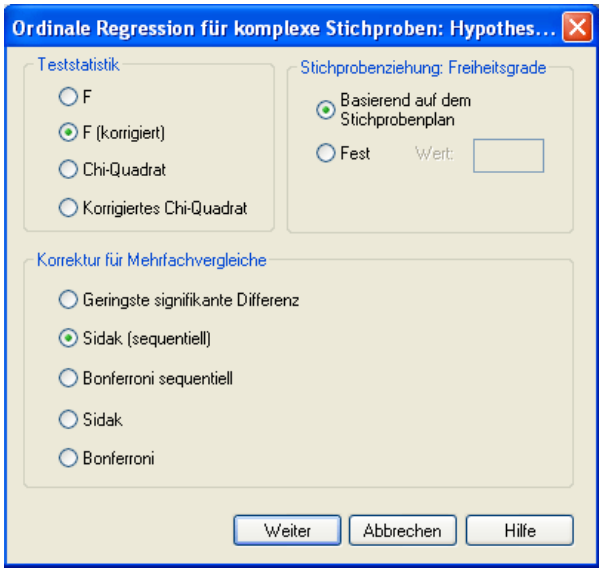
- Wählen Sie *The legislature should enact a gas tax* (Der Gesetzgeber sollte eine Kraftstoffsteuer einführen) als abhängige Variable.
- Wählen Sie *Age category* (Alterskategorie) bis *Driving frequency* (Häufigkeit der Kraftfahrzeugnutzung) als Faktoren aus.
- Klicken Sie auf Statistik.

Abbildung 21-3
Dialogfeld "Ordinale Regression: Statistik"



- Wählen Sie in der Gruppe "Anpassungsgüte des Modells" die Option Klassifikationsmatrix aus.
- Wählen Sie in der Gruppe "Parameter" die Optionen Schätzer, Exponentialfunktion des Schätzers, Standardfehler, Konfidenzintervall und Effekt des Stichprobenplans aus.
- Wählen Sie Wald-Test auf gleiche Neigungen und Parameterschätzer für verallgemeinertes Modell (Neigungen ungleich) aus.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Ordinale Regression für komplexe Stichproben" auf Hypothesentests.

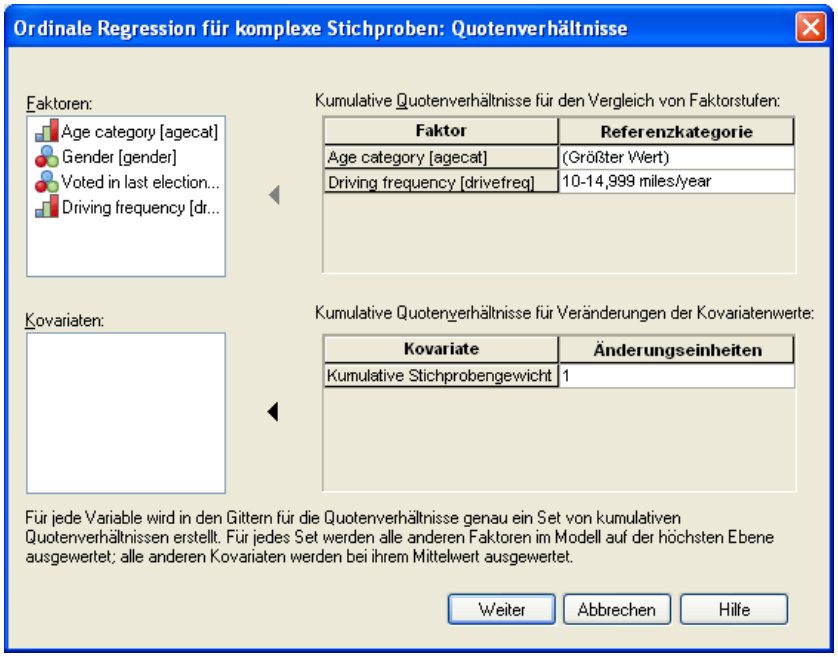
Abbildung 21-4
Dialogfeld "Hypothesentests"



Selbst bei einer relativ kleinen Anzahl an Einflussvariablen und Antwortkategorien kann die Wald*F*-Test-Statistik für den Parallelitätstest für Linien unschätzbar sein.

- Wählen Sie in der Gruppe "Teststatistik" die Option F (korrigiert) aus.
- Wählen Sie Sidak (sequentiell) als Anpassungsmethode für Mehrfachvergleiche aus.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Ordinale Regression für komplexe Stichproben" auf Quotenverhältnisse.

Abbildung 21-5
Dialogfeld "Ordinale Regression: Quotenverhältnisse"


Ordinale Regression für komplexe Stichproben: Quotenverhältnisse
Faktoren:
Age category [agecat]
Gender [gender]
Voted in last election...
Driving frequency [dr...
Kumulative Quotenverhältnisse für den Vergleich von Faktorstufen:
Faktor
Referenzkategorie
Age category [agecat]
(Größter Wert)
Driving frequency [drivefreq]
10-14,999 miles/year
Kovariaten:
Kumulative Quotenverhältnisse für Veränderungen der Kovariatenwerte:
Kovariate
Änderungseinheiten
Kumulative Stichprobengewicht
1
Für jede Variable wird in den Gittern für die Quotenverhältnisse genau ein Set von kumulativen Quotenverhältnissen erstellt. Für jedes Set werden alle anderen Faktoren im Modell auf der höchsten Ebene ausgewertet; alle anderen Kovariaten werden bei ihrem Mittelwert ausgewertet.
Weiter
Abbrechen
Hilfe

- Wählen Sie aus, dass für *Age category* (Alterskategorie) und *Driving frequency* (Häufigkeit der Kraftfahrzeugnutzung) kumulative Quotenverhältnisse berechnet werden sollen.
- Wählen Sie als Referenzkategorie für *Driving frequency* (Häufigkeit der Kraftfahrzeugnutzung) den Wert 10-14,999 miles/year (10.000 bis 14.999 Meilen/Jahr), eine typischere jährliche Fahrleistung als der größte Wert.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Ordinale Regression für komplexe Stichproben" auf OK.

Werte für Pseudo-R-Quadrat

Abbildung 21-6
Werte für Pseudo-R-Quadrat

Cox und Snell	,179
Nagelkerke	,191
McFadden	,071

Abhängige Variable: The legislature should enact a gas tax (Aufsteigend)
Modell: (Schwellenwert), agecat, gender, votelast, drivefreq
Verknüpfungsfunktion: Logit

Im linearen Regressionsmodell fasst das Bestimmtheitsmaß, R^2, den Anteil der Varianz in der abhängigen Variable, das mit den (unabhängigen) Einflussvariablen zu tun hat, zusammen. Dabei deuten größere R^2-Werte darauf hin, dass ein größerer Anteil der Varianz durch das Modell erklärt wird, bis zu einem Maximalwert von 1. Bei Regressionsmodellen mit einer kategorialen

abhängigen Variablen kann keine einzelne R^2-Statistik berechnet werden, die alle Merkmale von R^2 im linearen Regressionsmodell aufweist. Daher werden stattdessen diese Näherungen berechnet. Folgende Verfahren werden verwendet, um das Bestimmtheitsmaß abzuschätzen.

- R^2 nach Cox und Snell (Cox als auch Snell, 1989) beruht auf der Log-Likelihood für das Modell im Vergleich mit der Log-Likelihood für ein Grundlinienmodell. Bei kategorialen Ergebnissen hat es jedoch einen theoretischen Maximalwert von weniger als 1, sogar für ein "perfektes" Modell.
- R^2 nach Nagelkerke (Nagelkerke, 1991) ist eine korrigierte Version des *R*-Quadrats nach Cox & Snell, bei dem die Skala der Statistik so angepasst wird, dass sie den vollständigen Bereich von 0 bis 1 abdeckt.
- R^2 nach McFadden (McFadden, 1974) ist eine weitere Version, die auf den Log-Likelihood-Kernels für das Modell mit ausschließlich konstanten Termen und das vollständige geschätzte Modell beruht.

Was einen "guten" R^2-Wert ausmacht, hängt von den verschiedenen Anwendungsbereichen ab. Diese Statistiken können zwar auch für sich genommen bereits Schlüsse erlauben, sie sind jedoch am sinnvollsten, wenn es um den Vergleich von konkurrierenden Modellen für dieselben Daten geht. Das Modell mit der größten R^2-Statistik ist nach diesem Maßstab am "besten".

Tests der Modelleffekte

Abbildung 21-7
Tests der Modelleffekte

Quelle	Freiheits-grade 1	Freiheits-grade 2	Wald-F (korrigiert)	Sig.	Sidak-Sig. (sequentiell)
agecat	2,283	31,966	6,215	,004	,003
gender	1,000	14,000	,046	,834	,834
votelast	1,000	14,000	,076	,787	,787
drivefreq	3,785	52,987	228,015	,000	,000

Abhängige Variable: The legislature should enact a gas tax (Aufsteigend)
Modell: (Schwellenwert), agecat, gender, votelast, drivefreq
Verknüpfungsfunktion: Logit

Jeder Term im Modell wird daraufhin getestet, ob sein Effekt gleich 0 ist. Terme mit Signifikanzwerten von weniger als 0,05 weisen einen erkennbaren Effekt auf. Daher tragen *agecat* und *drivefreq* zum Modell bei, während die anderen Haupteffekte keinen Beitrag leisten. In einer weiteren Analyse der Daten könnten Sie die Entfernung von *gender* und *votelast* aus dem Modell in Erwägung ziehen.

Parameterschätzer

Die Tabelle der Parameterschätzer fasst den Effekt der einzelnen Einflussvariablen zusammen. Aufgrund des Charakters der Link-Funktion ist die Interpretation der Koeffizienten in diesem Modell zwar schwierig, die Vorzeichen der Koeffizienten für Kovariaten und die relativen Werte der Koeffizienten für Faktorstufen können jedoch wichtige Einblicke in die Effekte der Einflussvariablen im Modell bieten.

- Bei Kovariaten weisen positive (negative) Koeffizienten auf positive (inverse) Beziehungen zwischen Einflussvariablen und Ergebnis hin. Ein steigender Wert einer Kovariaten mit einem positiven Koeffizienten entspricht einer steigenden Wahrschreinlichkeit dafür, dass sie sich in einer der "höheren" Kategorien für das kumulative Ergebnis befindet.
- Bei Faktoren weist eine Faktorstufe mit einem größeren Koeffizienten auf eine größere Wahrscheinlichkeit dafür hin, dass sich der Faktor in einer der "höheren" Kategorien für das kumulative Ergebnis befindet. Das Vorzeichen eines Koeffizienten für eine Faktorstufe hängt von dem Effekt der betreffenden Faktorstufe in Bezug zur Referenzkategorie ab.

Abbildung 21-8
Parameterschätzer

Parameter		B	Standardfehler	95%-Konfidenzintervall Untere Grenze	95%-Konfidenzintervall Obere Grenze	Effekt des Stichprobenplans	Exp(B)	95% Konfidenzintervall für Exp(B) Untere Grenze	95% Konfidenzintervall für Exp(B) Obere Grenze
Schwellenwert	[opinion_gastax=4]	-3,343	,104	-3,566	-3,120	1,132	,035	,028	,044
	[opinion_gastax=3]	-1,910	,098	-2,120	-1,700	1,058	,148	,120	,183
	[opinion_gastax=2]	-,674	,090	-,866	-,482	,915	,510	,421	,618
Regression	[agecat=1]	-,324	,079	-,494	-,154	1,793	,723	,610	,858
	[agecat=2]	-,138	,054	-,255	-,022	1,158	,871	,775	,978
	[agecat=3]	-,085	,076	-,257	,088	2,206	,909	,773	1,070
	[agecat=4]	,000[a]	.	.	.	.	1,000	.	.
	[gender=0]	-,008	,035	-,084	,068	,949	,992	,920	1,071
	[gender=1]	,000[a]	.	.	.	.	1,000	.	.
	[votelast=0]	-,011	,039	-,095	,073	1,103	,989	,909	1,076
	[votelast=1]	,000[a]	.	.	.	.	1,000	.	.
	[drivefreq=1]	-3,751	,153	-4,079	-3,423	1,117	,023	,017	,033
	[drivefreq=2]	-3,003	,116	-3,251	-2,755	1,226	,050	,039	,064
	[drivefreq=3]	-2,295	,114	-2,540	-2,050	1,585	,101	,079	,129
	[drivefreq=4]	-1,570	,092	-1,769	-1,372	1,078	,208	,171	,254
	[drivefreq=5]	-,812	,089	-1,003	-,621	,941	,444	,367	,537
	[drivefreq=6]	,000[a]	.	.	.	.	1,000	.	.

Abhängige Variable: The legislature should enact a gas tax (Absteigend)
Modell: (Schwellenwert), agecat, gender, votelast, drivefreq
Verknüpfungsfunktion: Logit

a. Auf 0 gesetzt, da dieser Parameter redundant ist.

Auf der Grundlage der Parameterschätzer sind folgende Interpretationen möglich:

- Die Personen in den niedrigeren Alterskategorien zeigen größere Unterstützung für die Gesetzesvorlage als die Personen in der höchsten Alterskategorie.
- Die Personen, die weniger häufig ein Kraftfahrzeug nutzen, zeigen größere Unterstützung für die Gesetzesvorlage als Personen, die häufiger ein Kraftfahrzeug nutzen.
- Die Koeffizienten für die Variablen *gender* und *votelast* sind nicht nur nicht statistisch signifikant, sondern auch klein gegenüber den anderen Koeffizienten.

Die Effekte des Stichprobenplans weisen darauf hin, dass einige der für diese Parameterschätzer berechneten Standardfehler größer (und andere dagegen kleiner) sind als diejenigen, die man bei Verwendung einer einfachen Zufallsstichprobe erhalten würde. Es ist von entscheidender Bedeutung, die Informationen über den Stichprobenplan in die Analyse aufzunehmen, da Sie anderenfalls beispielsweise schließen könnten, dass der Koeffizient für die 3. Stufe von *Age category* (Alterskategorie), *[agecat=3]*, sich signifikant von 0 unterscheidet.

Klassifikation

Abbildung 21-9
Informationen zu kategorialen Variablen

		Gewichtete Anzahl	Prozent (gewichtet)
The legislature should enact a gas tax[a]	Strongly agree	25132,955	21,3%
	Agree	32261,425	27,3%
	Disagree	29477,417	24,9%
	Strongly disagree	31314,203	26,5%
Age category	18-30	20509,504	17,4%
	31-45	35380,506	29,9%
	46-60	34865,792	29,5%
	>60	27430,198	23,2%
Gender	Male	61424,547	52,0%
	Female	56761,453	48,0%
Voted in last election	No	70607,216	59,7%
	Yes	47578,784	40,3%
Driving frequency	Do not own car	3437,137	2,9%
	<10,000 miles/year	10816,349	9,2%
	10-14,999 miles/year	32539,364	27,5%
	15-19,999 miles/year	39179,814	33,2%
	20-29,999 miles/year	25617,804	21,7%
	>=30,000 miles/year	6595,532	5,6%
Umfang der Grundgesamtheit		118186,000	100,0%

a. Die Werte der abhängigen Variablen werden in absteigender Reihenfolge sortiert.

Mit den beobachteten Daten würde das “Nullmodell” (d. h. ein Modell ohne Einflussvariablen) alle Kunden in die Modalgruppe *Agree* (Stimme zu) einordnen. Das Nullmodell wäre also in 27,3 % der Fälle richtig.

Abbildung 21-10
Klassifikationsmatrix

	Vorhergesagt				
Beobachtet	Strongly agree	Agree	Disagree	Strongly disagree	Prozent korrekt
Strongly agree	7067,567	12130,814	3875,825	2058,750	28,1%
Agree	4271,234	14464,286	7320,767	6205,137	44,8%
Disagree	2024,816	11703,368	7108,487	8640,746	24,1%
Strongly disagree	889,869	8169,109	6946,522	15308,703	48,9%
Prozent (insgesamt)	12,1%	39,3%	21,4%	27,3%	37,2%

Abhängige Variable: The legislature should enact a gas tax (Absteigend)
Modell: (Schwellenwert), agecat, gender, votelast, drivefreq
Verknüpfungsfunktion: Logit

Die Klassifikationsmatrix zeigt die praktischen Ergebnisse der Verwendung des Modells. Für jeden Fall ist die vorhergesagte Antwortkategorie die Kategorie mit der höchsten vom Modell vorhergesagten Wahrscheinlichkeit. Die Fälle werden nach der *endgültigen Stichprobengewichtung* gewichtet, sodass die Klassifikationsmatrix die erwartete Modellleistung in der Grundgesamtheit wiedergibt.

- Die Zellen auf der Diagonale stellen korrekte Vorhersagen dar.
- Die Zellen abseits der Diagonale stellen falsche Vorhersagen dar.

Das Modell klassifiziert weitere 9,9 %, als 37,2 % der Klasse korrekt. Insbesondere bietet das Modell eine erheblich bessere Leistung bei der Klassifikation der Personen, die *Agree* (Stimme zu) oder *Strongly disagree* (Stimme ganz und gar nicht zu) ausgewählt haben, und ein wenig schlechter bei den Personen, die sich für *Disagree* (Stimme nicht zu) entschieden haben.

Quotenverhältnisse

Kumulative Quoten (Odds) sind definiert als Verhältnis zwischen der Wahrscheinlichkeit, dass die abhängige Variable einen Wert kleiner oder gleich einer bestimmten Antwortkategorie annimmt, und der Wahrscheinlichkeit, das sie einen Wert annimmt, der größer als die Antwortkategorie ist. Das **kumulative Quotenverhältnis** ist das Verhältnis der kumulativen Odds für verschiedene Einflusswerte und ist eng mit den potenzierten Parameterschätzern verwandt. Interessanterweise ist das kumulative Quotenverhältnis selbst nicht von der Antwortkategorie abhängig.

Abbildung 21-11
Kumulative Quotenverhältnisse für "Age category" (Alterskategorie).

		Kumulatives Quotenverhältnis	95%-Konfidenzintervall		Effekt des Stichprobenplans	Wurzel aus dem Effekt
			Untere Grenze	Obere Grenze		
Age category	18-30 vs. >60	1,383	1,166	1,639	1,793	1,339
	31-45 vs. >60	1,148	1,022	1,290	1,158	1,076
	46-60 vs. >60	1,100	,935	1,294	2,206	1,485

Abhängige Variable: The legislature should enact a gas tax (Absteigend)
Modell: (Schwellenwert), agecat, gender, votelast, drivefreq
Verknüpfungsfunktion: Logit

a. Die bei der Berechnung verwendeten Faktoren und Kovariaten werden auf folgende Werte festgesetzt: Age category=>60; Gender=Female; Voted in last election=Yes; Driving frequency=>=30,000 miles/year

In dieser Tabelle werden kumulative Quotenverhältnisse für die Faktorstufen von *Age category* (Alterskategorie) angezeigt. Bei den ausgegebenen Werten handelt es sich um die Verhältnisse der kumulativen Odds für *18–30* bis *46–60*, im Vergleich zu den kumulierten Odds für *>60*. Das Quotenverhältnis (Odds-Verhältnis) von 1,383 in der ersten Tabellenzeile bedeutet, dass die kumulativen Odds für eine Person im Alter von 18–30 das 1,383-Fache der kumualativen Odds für eine Person betragen, die älter als 60 ist. Beachten Sie: Da *Age category* (Alterskategorie) an keinem Wechselwirkungsterm beteiligt ist, handelt es sich bei den Quotenverhältnissen einfach um die Verhältnisse der potenzierten Parameterschätzer. So hat beispielsweise das kumulative Quotenverhältnis für *18–30* gegenüber *>60* den Wert 1,00/0,723 = 1,383.

Abbildung 21-12
Quotenverhältnisse für die Häufigkeit der Kraftfahrzeugnutzung

		Kumulatives Quotenverhältnis	95%-Konfidenzintervall		Effekt des Stichprobenplans	Wurzel aus dem Effekt
			Untere Grenze	Obere Grenze		
Driving frequency	Do not own car vs. 10-14,999 miles/year	4,288	2,878	6,390	2,345	1,531
	<10,000 miles/year vs. 10-14,999 miles/year	2,030	1,656	2,488	1,838	1,356
	15-19,999 miles/year vs. 10-14,999 miles/year	,484	,430	,546	1,450	1,204
	20-29,999 miles/year vs. 10-14,999 miles/year	,227	,193	,267	2,095	1,448
	>=30,000 miles/year vs. 10-14,999 miles/year	,101	,079	,129	1,585	1,259

Abhängige Variable: The legislature should enact a gas tax (Absteigend)
Modell: (Schwellenwert), agecat, gender, votelast, drivefreq
Verknüpfungsfunktion: Logit

a. Die bei der Berechnung verwendeten Faktoren und Kovariaten werden auf folgende Werte festgesetzt: Age category=>60; Gender=Female; Voted in last election=Yes; Driving frequency=>=30,000 miles/year

In dieser Tabelle werden die kumulativen Quotenverhältnisse für die Faktorstufen von *Driving frequency* "Häufigkeit der Kraftfahrzeugnutzung" angezeigt, wobei *10–14,999 miles/year* (10.000 bis 14.999 Meilen/Jahr) als Referenzkategorie verwendet wird. Da *Driving frequency* "Häufigkeit der Kraftfahrzeugnutzung" an keinem Wechselwirkungsterm beteiligt ist, handelt es sich bei den Quotenverhältnissen einfach um die Verhältnisse der potenzierten Parameterschätzer. So beträgt beispielsweise das kumulative Quotenverhältnis für *20–29,999 miles/year* (20.000 bis 29.999 Meilen/Jahr) gegenüber *10–14,999 miles/year* (10.000 bis 14.999 Meilen/Jahr) 0,101/0,444 = 0,227.

Verallgemeinertes kumulatives Modell

Abbildung 21-13
Parallelitätstest für Linien

Freiheitsgrade 1	Freiheitsgrade 2	Wald-F (korrigiert)	Sig.	Sidak-Sig. (sequentiell)
8,769	122,767	1,894	,061	,392

Abhängige Variable: The legislature should enact a gas tax (Absteigend)
Modell: (Schwellenwert), agecat, gender, votelast, drivefreq
Verknüpfungsfunktion: Logit

Der Parallelitätstest für Linien kann Ihnen helfen einzuschätzen, ob die Annahme, dass die Parameter für alle Antwortkategorien gleich sind, plausibel ist. Bei diesem Test wird das geschätzte Modell mit einem einzigen Satz von Koeffizienten für alle Kategorien mit einem generalisierten Modell mit einem separaten Satz von Koeffizienten für die einzelnen Kategorien verglichen.

Der Wald-*F*-Test ist ein Omnibus-Test der Kontrastmatrix für die Annahme paralleler Linien, der asymptotisch korrekte *p*-Werte bietet; für kleine bis mittelgroße Stichproben ist die korrigierteWald-*F*-Statistik gut geeignet. Der Signifikanzwert liegt nahe an 0,05, was darauf hindeutet, dass das verallgemeinerte Modell eine Verbesserung der Modellanpassung bieten könnte; der korrigierte sequenzielle Sidak-Test gibt jedoch einen hinreichend hohen

Signifikanzwert aus (0,392), sodass insgesamt kein eindeutiger Beweis für die Zurückweisung der Annahme paralleler Linien vorliegt. Der sequenzielle Sidak-Test beginnt mit Wald-Tests für Einzelkontraste, um einen Gesamt-*p*-Wert zu berechnen. Diese Ergebnisse sollten mit dem Ergebnis des Omnibus-Wald-Tests vergleichbar sein. Die Tatsache, dass sie im vorliegenden Beispiel so unterschiedlich sind, überrascht. Dies könnte jedoch daran liegen, dass im Test recht viele Kontraste und Freiheitsgrade für einen relativ kleinen Stichprobenplan vorliegen.

Abbildung 21-14
Parameterschätzer für verallgemeinertes kumulatives Modell (teilweise gezeigt).

The legislature should enact a gas tax	Parameter	B	Standardfehler	95%-Konfidenzintervall Untere	Obere
Strongly agree	(Threshold)	-3,681	,221	-4,155	-3,207
	[agecat=1]	-,320	,096	-,525	-,115
	[agecat=2]	-,075	,071	-,227	,077
	[agecat=3]	-,022	,073	-,180	,135
	[agecat=4]	,000[a]	.		
	[gender=0]	-,082	,054	-,197	,033
	[gender=1]	,000[a]	.		
	[votelast=0]	,008	,052	-,104	,120
	[votelast=1]	,000[a]	.		
	[drivefreq=1]	-4,096	,267	-4,669	-3,523
	[drivefreq=2]	-3,367	,237	-3,876	-2,857
	[drivefreq=3]	-2,678	,224	-3,158	-2,199
	[drivefreq=4]	-1,928	,213	-2,384	-1,471
	[drivefreq=5]	-1,015	,252	-1,555	-,476
	[drivefreq=6]	,000[a]	.		
Agree	(Threshold)	-1,963	,153	-2,291	-1,635
	[agecat=1]	-,385	,095	-,587	-,182
	[agecat=2]	-,130	,069	-,279	,018
	[agecat=3]	-,139	,101	-,356	,077
	[agecat=4]	,000[a]	.		
	[gender=0]	-,004	,040	-,090	,082
	[gender=1]	,000[a]	.		
	[votelast=0]	,009	,059	-,117	,135
	[votelast=1]	,000[a]	.		
	[drivefreq=1]	-3,867	,318	-4,549	-3,185
	[drivefreq=2]	-3,005	,175	-3,380	-2,630
	[drivefreq=3]	-2,290	,187	-2,691	-1,888
	[drivefreq=4]	-1,633	,166	-1,988	-1,278
	[drivefreq=5]	-,909	,137	-1,204	-,615
	[drivefreq=6]	,000[a]	.		

Darüber hinaus scheinen sich die geschätzten Werte der Koeffizienten des verallgemeinerten Modells nicht stark von den Schätzern unter der Annahme paralleler Linien zu unterscheiden.

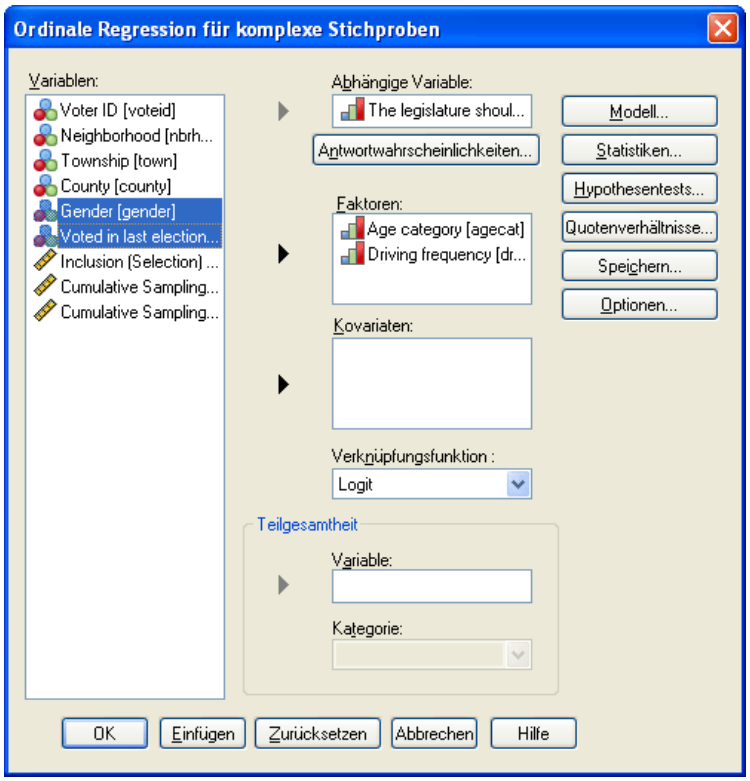
Verwerfen nichtsignifikanter Einflussvariablen

Die Tests von Modelleffekten haben gezeigt, dass die Modellkoeffizienten für *Gender* (Geschlecht) und *Voted in last election* (An der letzten Wahl teilgenommen) nicht statistisch signifikant von 0 abweichen.

- Um ein reduziertes Modell zu erstellen, rufen Sie das Dialogfeld “Ordinale Regression für komplexe Stichproben” erneut auf.

▶ Klicken Sie im Dialogfeld “Plan” auf Weiter.

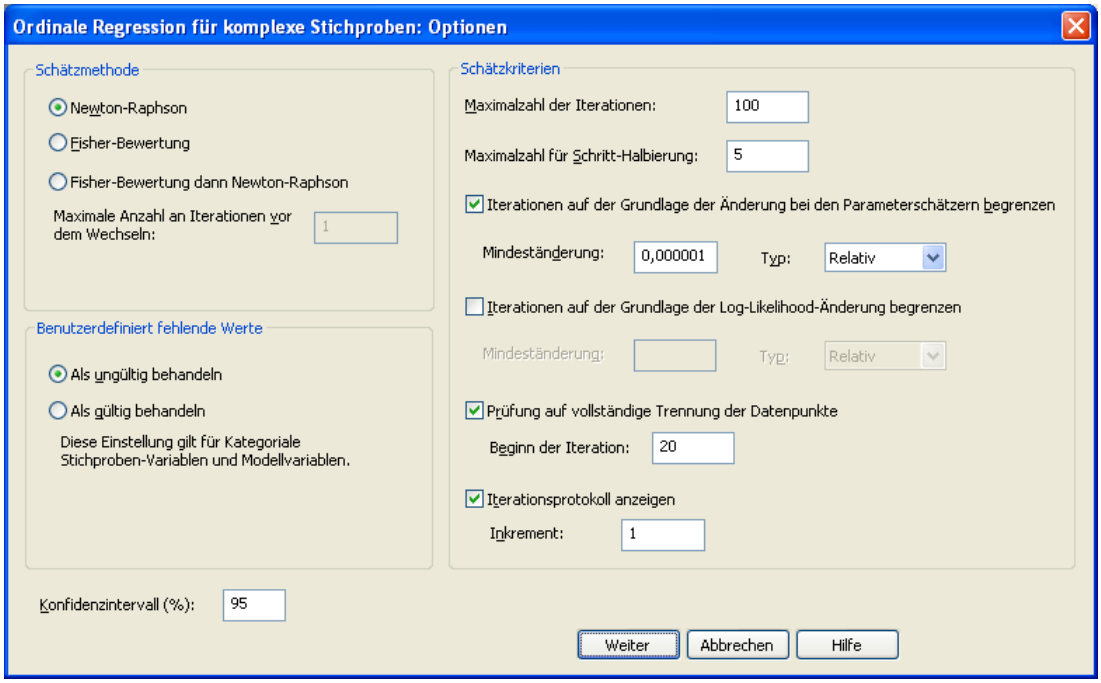
Abbildung 21-15
Dialogfeld “Ordinale Regression”



▶ Heben Sie die Auswahl von *Gender* (Geschlecht) und *Voted in last election* (An der letzten Wahl teilgenommen) als Faktoren auf.

▶ Klicken Sie auf Optionen.

Abbildung 21-16
Dialogfeld "Ordinale Regression: Optionen"



- Wählen Sie die Option Iterationsprotokoll anzeigen.

  Das Iterationsprotokoll dient zur Diagnostizierung der Probleme, auf die der Schätzalgorithmus stößt.

- Klicken Sie auf Weiter.

- Klicken Sie im Dialogfeld "Ordinale Regression für komplexe Stichproben" auf OK.

Warnungen

Abbildung 21-17
Warnungen für das reduzierte Modell

Der Log-Likelihood-Wert kann nach der maximalen Anzahl der Schritte in der Schritt-Halbierungs-Methode nicht weiter erhöht werden.

Die Prozedur CSORDINAL wird trotz der oben stehenden Warnung(en) fortgesetzt. Die angezeigten nachfolgenden Ergebnisse beruhen auf der letzten Iteration. Die Gültigkeit der Anpassungsgüte des Modells ist ungewiss.

Die folgende Meldung gilt für das verallgemeinerte kumulative Modell.

Der Log-Likelihood-Wert kann nach der maximalen Anzahl der Schritte in der Schritt-Halbierungs-Methode nicht weiter erhöht werden.

In den Warnungen wird angegeben, dass die Schätzung des reduzierten Modells endete, bevor die Parameterschätzer Konvergenz erreichten, da die Log-Likelihood nicht durch eine Änderung bzw. einen "Schritt" in den aktuellen Werten der Parameterschätzer erhöht werden konnte.

Abbildung 21-18
Warnungen für das reduzierte Modell

Iteration[b]	N Schritt-Halbierungen	Pseudo -2 Log-Likelihood	Schwellenwert			Regression							
			[opinion_gastax=1]	[opinion_gastax=2]	[opinion_gastax=3]	[agecat=1]	[agecat=2]	[agecat=3]	[drivefreq=1]	[drivefreq=2]	[drivefreq=3]	[drivefreq=4]	[drivefreq=5]
0	0	326640,341	-1,309	-,058	1,020	,000	,000	,000	,000	,000	,000	,000	,000
1	0	303567,549	-3,242	-1,881	-,704	-,323	-,137	-,094	-3,841	-2,970	-2,248	-1,563	-,835
2	0	303336,336	-3,327	-1,897	-,664	-,325	-,139	-,095	-3,740	-2,998	-2,291	-1,568	-,811
3	0	303335,933	-3,333	-1,900	-,664	-,326	-,139	-,096	-3,750	-3,003	-2,295	-1,570	-,812
4	0	303335,933	-3,333	-1,900	-,664	-,326	-,139	-,096	-3,750	-3,003	-2,295	-1,570	-,812
5[a]	5	303335,933	-3,333	-1,900	-,664	-,326	-,139	-,096	-3,750	-3,003	-2,295	-1,570	-,812

Redundante Parameter werden nicht angezeigt. Ihre Werte sind in allen Iterationen stets null.
Abhängige Variable: The legislature should enact a gas tax (Aufsteigend)
Modell: (Schwellenwert), agecat, drivefreq
Verknüpfungsfunktion: Logit

a. Der Log-Likelihood-Wert kann nach der maximalen Anzahl der Schritte in der Schritt-Halbierungs-Methode nicht weiter erhöht werden.

b. Zur Schätzung der Parameter wurde die Newton-Raphson-Methode verwendet.

Wenn Sie das Iterationsprotokoll betrachten, werden sie feststellen, dass die Änderungen in den Paramterschätzern bei den letzten paar Iterationen so gering sind, dass die Warnmeldung keinen Anlass zur Sorge darstellt.

Vergleichen von Modellen

Abbildung 21-19
Pseudo-R-Quadrat-Werte für das reduzierte Modell

Cox und Snell	,179
Nagelkerke	,191
McFadden	,071

Abhängige Variable: The legislature should enact a gas tax (Aufsteigend)
Modell: (Schwellenwert), agecat, gender, votelast, drivefreq
Verknüpfungsfunktion: Logit

Die R^2-Werte für das reduzierte Modell sind mit den Werten für das ursprüngliche Modell identisch. Dies spricht für das reduzierte Modell.

Abbildung 21-20
Klassifikationsmatrix für das reduzierte Modell

Beobachtet	Vorhergesagt				
	Strongly agree	Agree	Disagree	Strongly disagree	Prozent korrekt
Strongly agree	7067,567	12823,258	3183,380	2058,750	28,1%
Agree	4271,234	15684,090	6100,963	6205,137	48,6%
Disagree	2024,816	13157,809	5654,047	8640,746	19,2%
Strongly disagree	889,869	9226,578	5889,053	15308,703	48,9%
Prozent (insgesamt)	12,1%	43,1%	17,6%	27,3%	37,0%

Abhängige Variable: The legislature should enact a gas tax (Aufsteigend)
Modell: (Schwellenwert), agecat, drivefreq
Verknüpfungsfunktion: Logit

Die Klassifikationsmatrix verkompliziert die Sache ein wenig. Die Gesamtklassifizierungsquote von 37,0 % für das reduzierte Modell ist mit dem ursprünglichen Modell vergleichbar, was für das reduzierte Modell spricht. Das reduzierte Modell verlagert jedoch die vorhergesagte Antwort

von 3,8 % der Wäher von *Disagree* (Stimme nicht zu) zu *Agree* (Stimme zu). Den beobachteten Daten zufolge antwortete mehr als die Hälfte davon mit *Disagree* (Stimme nicht zu) bzw. *Strongly disagree* (Stimme ganz und gar nicht zu). Dies ist eine sehr wichtige Unterscheidung, die sorgfältiger Erwägung bedarf, bevor das reduzierte Modell gewählt wird.

Zusammenfassung

Sie haben mithilfe der ordinalen Regression für komplexe Stichproben konkurrierende Modelle für die Stärke der Unterstützung für den vorgeschlagenen Gesetzesentwurf auf der Grundlage der demografischen Struktur der Wähler konstruiert. Der Parallelitätstest für Linien zeigt, dass ein verallgemeinertes kumulatives Modell nicht erforderlich ist. Die Tests der Modelleffekte legen nahe, dass *Gender* (Geschlecht) und *Voted in last election* (An der letzten Wahl teilgenommen) aus dem Modell herausgenommen werden können und dass das reduzierte Modell hinsichtlich Pseudo-R^2 und Gesamtklassifizierungsquote im Vergleich zum ursprünglichen Modell sehr gut funktioniert. Das reduzierte Modell klassifiziert jedoch mehr Wähler hinsichtlich der Trennlinie *Agree* (Stimme zu)/*Disagree* (Stimme nicht zu) falsch, weshalb sich der Gesetzgeber vorerst für die Beibehaltung des ursprünglichen Modells entschieden hat.

Verwandte Prozeduren

Die Prozedur “Ordinale Regression für komplexe Stichproben” ist ein nützliches Tool für die Modellierung einer ordinalen Variablen, wenn die Fälle anhand eines Schemas für komplexe Stichproben gezogen wurden.

- Der Stichprobenassistent für komplexe Stichproben wird zur Angabe der Planspezifikationen für komplexe Stichproben und zum Ziehen von Stichproben verwendet. Die vom Stichprobenassistenten erstellte Stichprobenplan-Datei enthält einen Standard-Analyseplan und kann im Dialogfeld “Plan” angegeben werden, wenn die gezogene Stichprobe gemäß diesem Plan analysiert werden soll.
- Der Analysevorbereitungsassistent für komplexe Stichproben wird zur Angabe der Analysespezifikationen für eine bestehende komplexe Stichprobe verwendet. Die vom Stichprobenassistenten erstellte Analyseplan-Datei kann im Dialogfeld “Plan” angegeben werden, wenn Sie die Stichprobe gemäß diesem Plan analysieren.
- Die Prozedur Allgemeines lineares Modell für komplexe Stichproben ermöglicht die Modellierung einer metrischen Antwort (Responsevariablen).
- Die Prozedur Logistische Regression für komplexe Stichproben ermöglicht die Modellierung einer kategorialen Antwort (Responsevariablen).

Kapitel
22

Cox-Regression für komplexe Stichproben

Die Prozedur “Cox-Regression für komplexe Stichproben” besteht aus einer Überlebensanalyse für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden.

Verwenden einer zeitabhängigen Einflussvariablen in der Cox-Regression für komplexe Stichproben

Eine Strafverfolgungsbehörde ist hinsichtlich der Rückfallraten in ihrem Zuständigkeitsbereich unsicher. Eine der Messwerte der Rückfallrate ist die Zeit bis zur zweiten Festnahme von Straftätern. Die Behörde möchte die Zeit bis zur erneuten Festnahme mithilfe der Anwendung der Cox-Regression auf eine Stichprobe modellieren, die mit Methoden für komplexe Stichproben gezogen wurde, ist jedoch besorgt, dass die proportionale Hazard-Annahme für die einzelnen Alterskategorien ungültig ist.

Personen, die im Juni 2003 erstmals aus der Haft entlassen wurden, wurden aus per Stichprobenziehung ermittelten Polizeidirektionen ausgewählt und Ihr Fall wurde jeweils bis Ende Juni 2006 verfolgt. Die Stichprobe befindet sich in *recidivism_cs_sample.sav*. Der verwendete Stichprobenplan befindet sich in *recidivism_cs.csplan*. Da hier die PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*recidivism_cs_jointprob.sav*). Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Verwenden Sie die Cox-Regression für komplexe Stichproben, um die Gültigkeit der Annahme proportionaler Hazard-Raten zu bewerten und, falls angemessen, um ein Modell mit zeitabhängigen Einflussvariablen zu erstellen.

Vorbereitung der Daten

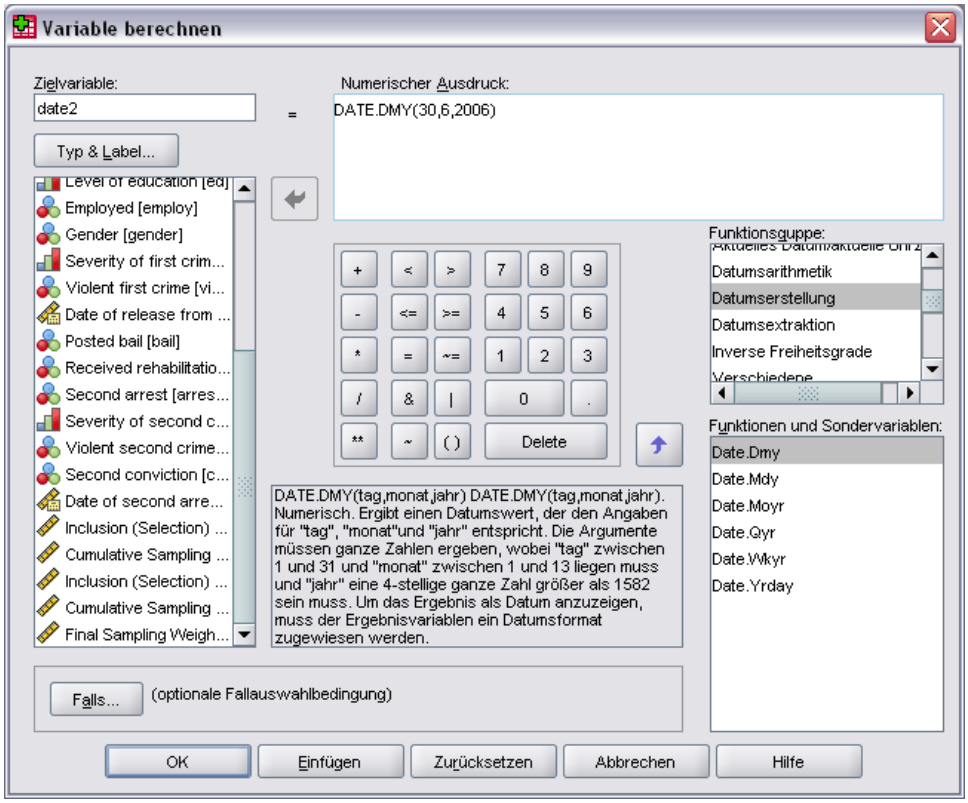
Das Daten-Set enthält die Daten der Freilassung nach der ersten Festnahme und die der zweiten Festnahme; da die Cox-Regression die Überlebenszeiten analysiert, müssen Sie die Zeitdauer zwischen diesen Daten berechnen.

Date of second arrest [date2] (Datum der zweiten Festnahme) enthält jedoch Fälle mit dem Wert “10/03/1582”, einem fehlenden Wert für Datumsvariablen. Hierbei handelt es sich um Personen, bei denen kein zweites Vergehen vorlag. Diese sollen auf jeden Fall als rechtszensierte Fälle in das Modell aufgenommen werden. Der Überwachungszeitraum endete am 30. Juni 2006, weshalb wir “10/03/1582” in “10/03/1582” umkodieren.

▶ Um diese Werte neu zu kodieren, wählen Sie die folgenden Menübefehle aus:

Transformieren
Variable berechnen...

Abbildung 22-1
Dialogfeld "Variable berechnen"

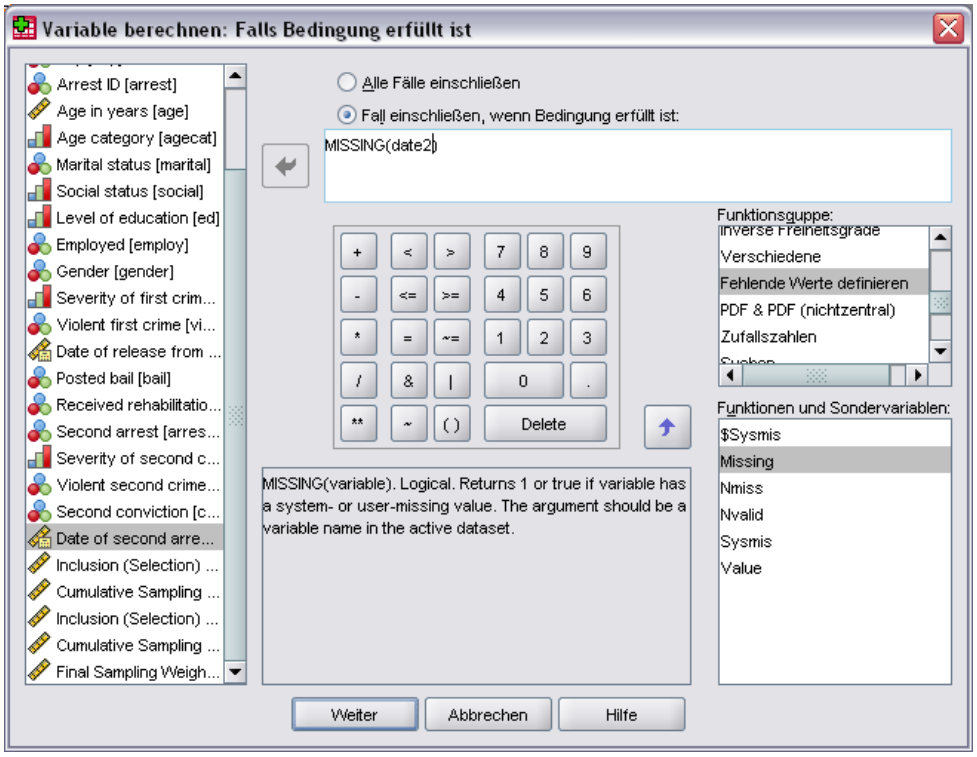


▶ Geben Sie date2 als Zielvariable ein.

▶ Geben Sie DATE.DMY(30,6,2006) als Ausdruck ein.

▶ Klicken Sie auf Falls.

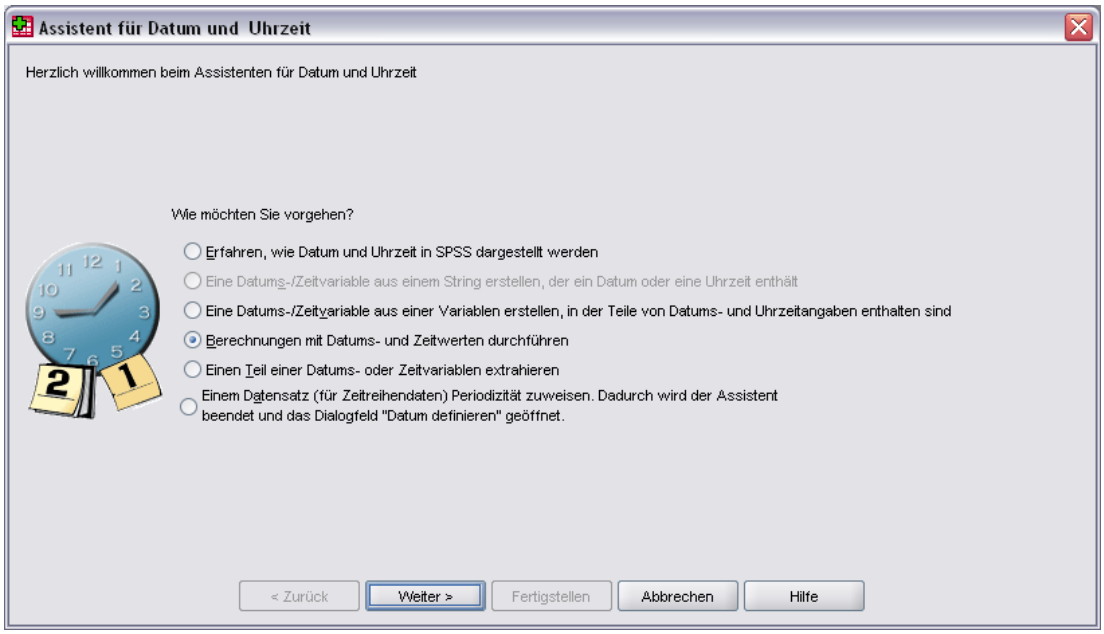
Abbildung 22-2
Dialogfeld "Variable berechnen: Falls Bedingung erfüllt ist"



- Wählen Sie Fall einschließen, wenn Bedingung erfüllt ist aus.
- Geben Sie MISSING(date2) als Ausdruck ein.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Variable berechnen" auf OK.
- Wählen Sie als Nächstes zur Berechnung des Zeitabstands zwischen erster und zweiter Festnahme folgende Optionen in den Menüs aus:

  Transformieren
  Assistent für Datum und Uhrzeit...

Abbildung 22-3
Assistent für Datum und Uhrzeit: Schritt "Willkommen"


Assistent für Datum und Uhrzeit
Herzlich willkommen beim Assistenten für Datum und Uhrzeit
Wie möchten Sie vorgehen?
Erfahren, wie Datum und Uhrzeit in SPSS dargestellt werden
Eine Datums-/Zeitvariable aus einem String erstellen, der ein Datum oder eine Uhrzeit enthält
Eine Datums-/Zeitvariable aus einer Variablen erstellen, in der Teile von Datums- und Uhrzeitangaben enthalten sind
Berechnungen mit Datums- und Zeitwerten durchführen
Einen Teil einer Datums- oder Zeitvariablen extrahieren
Einem Datensatz (für Zeitreihendaten) Periodizität zuweisen. Dadurch wird der Assistent beendet und das Dialogfeld "Datum definieren" geöffnet.
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

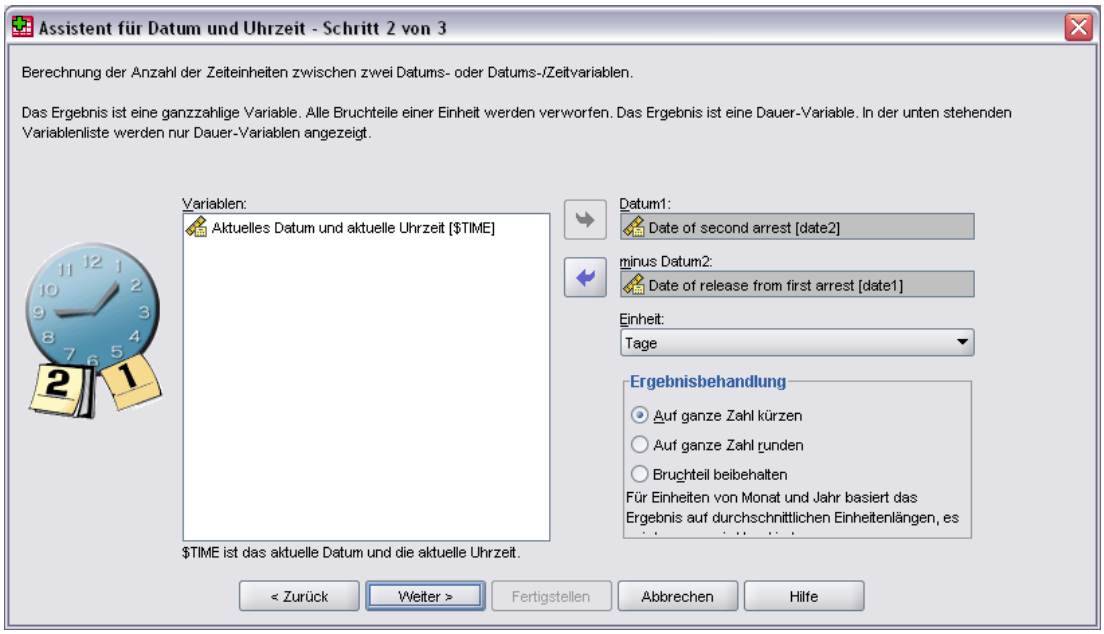
- Wählen Sie Berechnungen mit Datums- und Zeitwerten durchführen.
- Klicken Sie auf Weiter.

Abbildung 22-4
Assistent für Datum und Uhrzeit: Schritt "Durchführen von Berechnungen mit Datumswerten"

Assistent für Datum und Uhrzeit - Schritt 1 von 3
Durchführen von Berechnungen mit DatumswertenWählen Sie eine der folgenden Aufgaben aus und klicken Sie auf "Weiter"
Addieren bzw. Subtrahieren einer Dauer zu bzw. von einem Datum (z.B. Hinzufügen eines Monats zu einem Alter oder Addieren einer Zeitvariablen zu einer Datums-/Zeitvariablen)
Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten (z.B. Berechnen eines Alters in Jahren aus dem Geburtsdatum und einem weiteren Datum)
Subtrahieren zweier Werte für Dauer (z.B. Arbeitszeit - Anfahrtszeit)
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

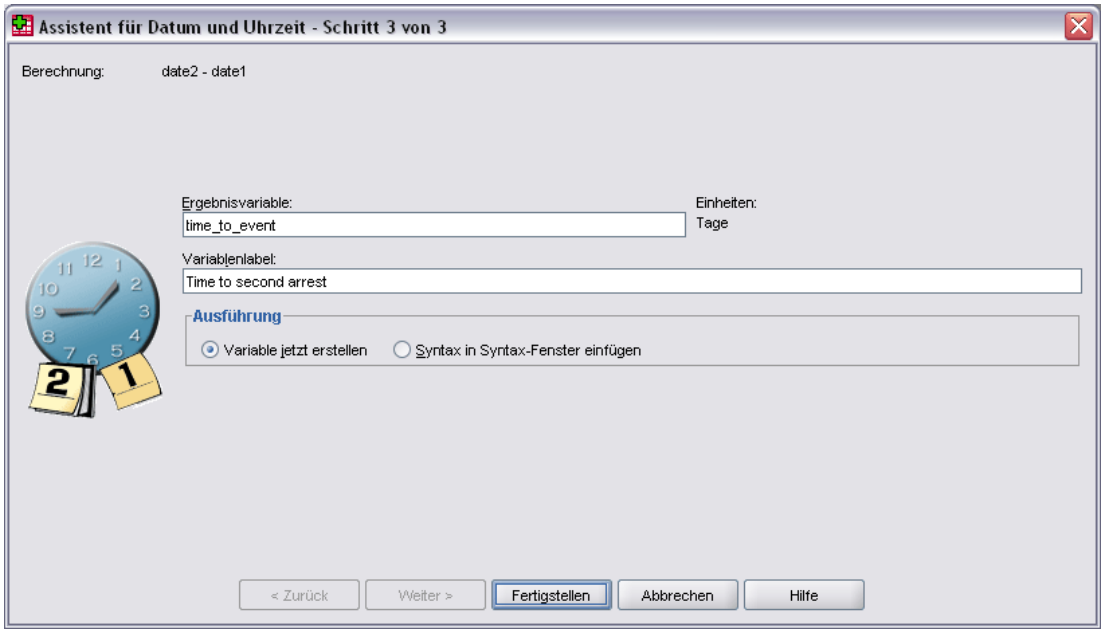
- Wählen Sie Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten.
- Klicken Sie auf Weiter.

Abbildung 22-5
Assistent für Datum und Uhrzeit: Schritt "Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten"


Assistent für Datum und Uhrzeit - Schritt 2 von 3
Berechnung der Anzahl der Zeiteinheiten zwischen zwei Datums- oder Datums-/Zeitvariablen.
Das Ergebnis ist eine ganzzahlige Variable. Alle Bruchteile einer Einheit werden verworfen. Das Ergebnis ist eine Dauer-Variable. In der unten stehenden Variablenliste werden nur Dauer-Variablen angezeigt.
Variablen:
Aktuelles Datum und aktuelle Uhrzeit [$TIME]
$TIME ist das aktuelle Datum und die aktuelle Uhrzeit.
Datum1:
Date of second arrest [date2]
minus Datum2:
Date of release from first arrest [date1]
Einheit:
Tage
Ergebnisbehandlung
Auf ganze Zahl kürzen
Auf ganze Zahl runden
Bruchteil beibehalten
Für Einheiten von Monat und Jahr basiert das Ergebnis auf durchschnittlichen Einheitenlängen, es
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

- Wählen Sie *Date of second arrest [date2]* (Datum der zweiten Festnahme) als erstes Datum aus.
- Wählen Sie *Date of release from first arrest [date1]* (Datum der Freilassung nach erster Festnahme) als das Datum aus, das vom ersten Datum subtrahiert werden soll.
- Wählen Sie Tage als Einheit aus.
- Klicken Sie auf Weiter.

Abbildung 22-6
Assistent für Datum und Uhrzeit: Schritt "Berechnung".



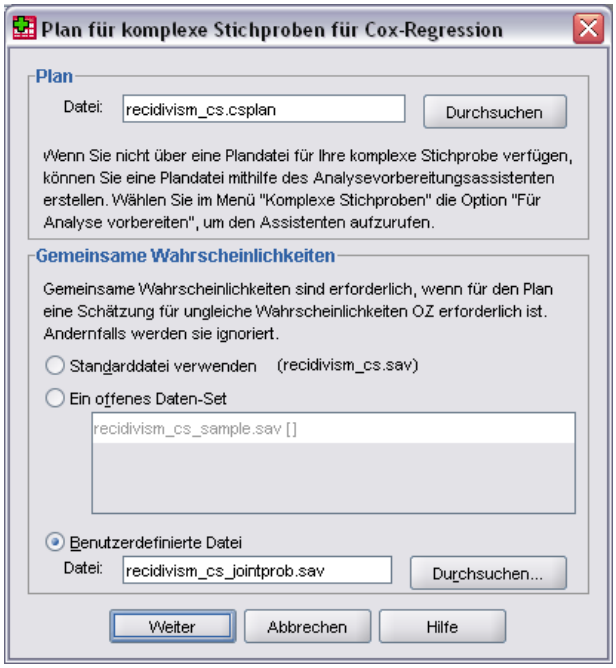
- Geben Sie *time_to_event* als Namen der Variablen ein, die für die Zeit zwischen den beiden Datumswerten steht.
- Geben Sie *Time to second arrest* (Zeit bis zur zweiten Festnahme) als Variablenlabel ein.
- Klicken Sie auf Fertig stellen.

Durchführung der Analyse

- Um eine Analyse der Art "Cox-Regression für komplexe Stichproben" durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

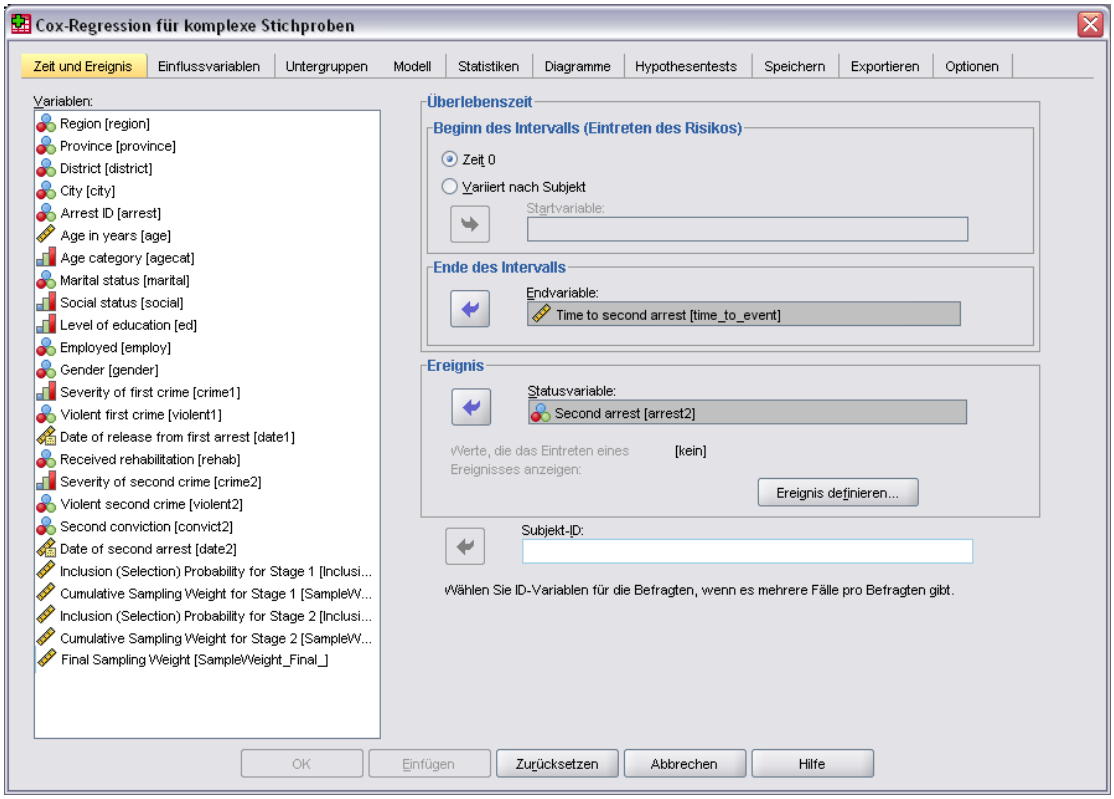
Analysieren
  Komplexe Stichproben
    Cox-Regression…

Abbildung 22-7
Dialogfeld "Cox-Regression für komplexe Stichproben"


Plan für komplexe Stichproben für Cox-Regression
Plan
Datei:
recidivism_cs.csplan
Durchsuchen
Wenn Sie nicht über eine Plandatei für Ihre komplexe Stichprobe verfügen, können Sie eine Plandatei mithilfe des Analysevorbereitungsassistenten erstellen. Wählen Sie im Menü "Komplexe Stichproben" die Option "Für Analyse vorbereiten", um den Assistenten aufzurufen.
Gemeinsame Wahrscheinlichkeiten
Gemeinsame Wahrscheinlichkeiten sind erforderlich, wenn für den Plan eine Schätzung für ungleiche Wahrscheinlichkeiten OZ erforderlich ist. Andernfalls werden sie ignoriert.
Standarddatei verwenden (recidivism_cs.sav)
Ein offenes Daten-Set
recidivism_cs_sample.sav []
Benutzerdefinierte Datei
Datei:
recidivism_cs_jointprob.sav
Durchsuchen...
Weiter
Abbrechen
Hilfe

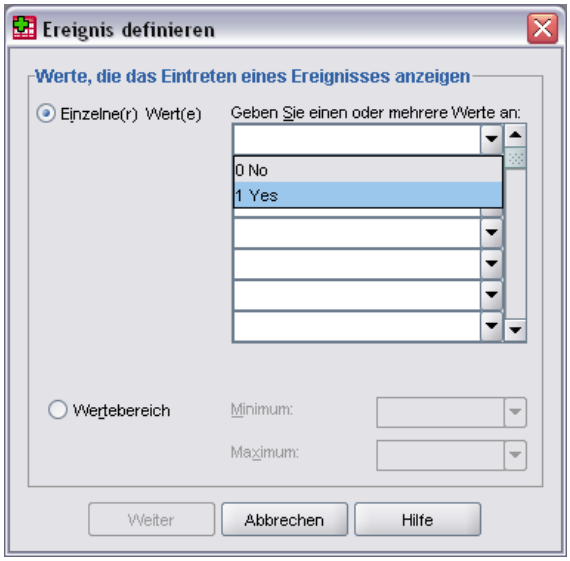
- Wechseln Sie zu dem Verzeichnis mit den Beispieldateien und wählen Sie *recidivism_cs.csplan* als Plandatei aus.
- Wählen Sie im Gruppenfeld "Gemeinsame Wahrscheinlichkeiten" die Option Benutzerdefinierte Datei und wählen Sie die Datei *recidivism_cs_jointprob.sav* aus.
- Klicken Sie auf Weiter.

Abbildung 22-8
Registerkarte "Zeit und Ereignis", Dialogfeld "Cox-Regression"



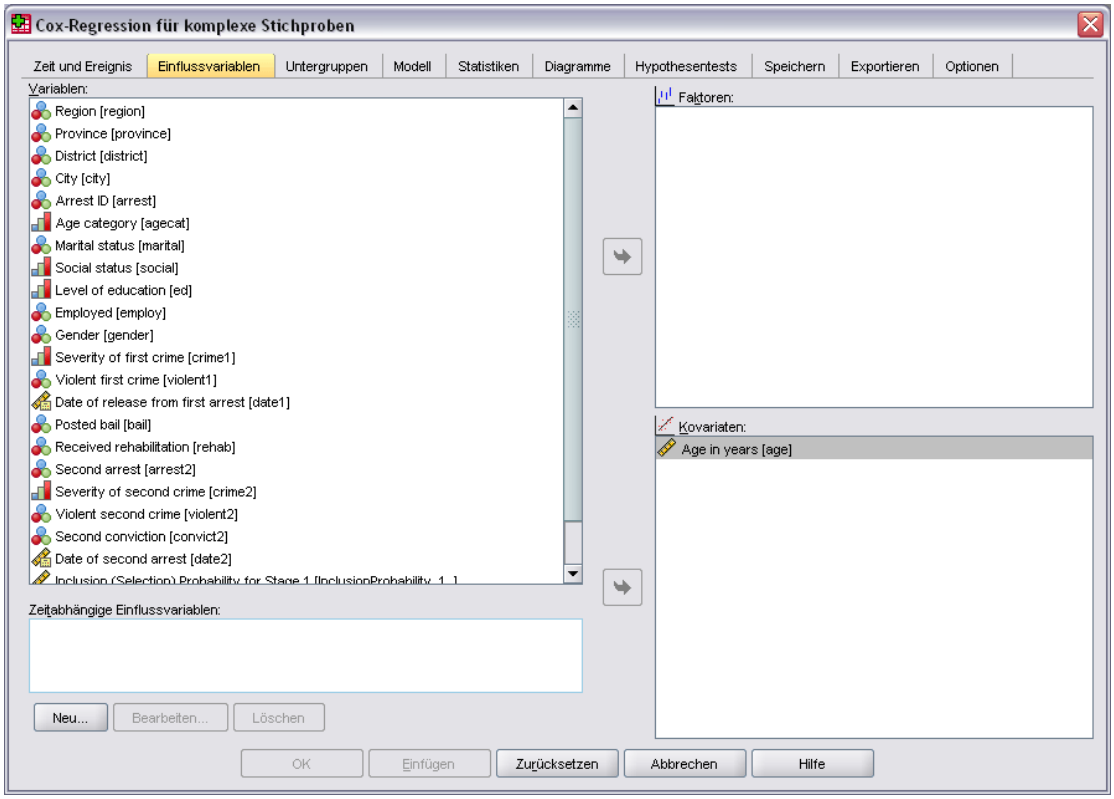
- Wählen Sie *Time to second arrest [time_to_event]* (Zeit bis zur zweiten Festnahme) als die Variable, die das Ende des Intervalls definiert.
- Wählen Sie *Second arrest [arrest2]* (Zweite Festnahme) als Variable, die definiert, ob das Ereignis eingetreten ist.
- Klicken Sie auf Ereignis definieren.

Abbildung 22-9
Dialogfeld "Ereignis definieren"



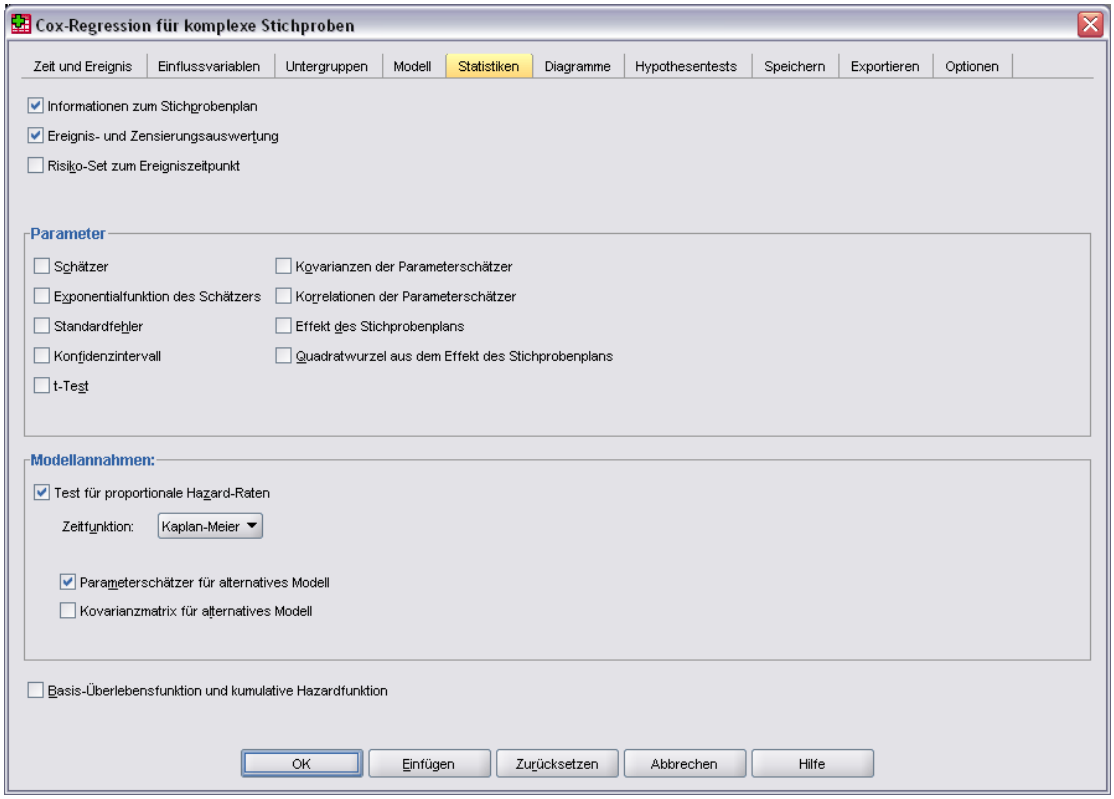
- Wählen Sie 1 Yes (Ja) als Wert, der anzeigt, dass das relevante Ereignis (erneute Festnahme) eingetreten ist.
- Klicken Sie auf Weiter.
- Klicken Sie auf die Registerkarte Einflussvariablen.

Abbildung 22-10
Dialogfeld "Cox-Regression", Registerkarte "Statistiken"



- Wählen Sie *Age in years [age]* (Alter in Jahren) als Kovariate aus.
- Klicken Sie auf die Registerkarte Statistik.

Abbildung 22-11
Dialogfeld "Cox-Regression", Registerkarte "Statistiken"



- Wählen Sie Test für proportionale Hazard-Raten aus und wählen Sie dann im Gruppenfeld "Modellannahmen" Log als Zeitfunktion aus.
- Wählen Sie Parameterschätzer für alternatives Modell.
- Klicken Sie auf OK.

Informationen zum Stichprobenplan

Abbildung 22-12
Informationen zum Stichprobenplan

		N
Gültig	Subjekte	5687
	Fälle	5687
	Ungültige Fälle	0
	Gesamtzahl der Fälle	5687
	Subjektgröße bei Grundgesamtheit	307583,898
	Schichten	4
	Einheiten	20
	Freiheitsgrade des Stichprobenplans	16

Diese Tabelle enthält Informationen zu dem Stichprobenplan der zur Schätzung des Modells gehört.

- Es gibt einen Fall pro Subjekt und alle 5.687 Fälle werden in der Analyse verwendet.
- Die Stichprobe stellt weniger als 2 % der gesamten geschätzten Grundgesamtheit dar.
- Der Stichprobenplan forderte 4 Schichten und 5 Einheiten pro Schicht für insgesamt 20 Einheiten in der ersten Stufe des Plans an. Die Freiheitsgrade des Stichprobenplans werden als 20−4=16 geschätzt.

Tests der Modelleffekte

Abbildung 22-13
Tests der Modelleffekte

Quelle	Freiheitsgrade 1	Freiheitsgrade 2	Wald-F	Sig.
age	1,000	16,000	504,787	1,580E-13

Überlebenszeitvariable: Time to second arrest
Ereignisstatusvariable: Second arrest = 1
Modell: age

im proportionalen Hazard-Modell liegt der Signifikanzwert für die Einflussvariable *age* (Alter) unter 0,05 und scheint somit einen Beitrag zu dem Modell zu leisten.

Test für proportionale Hazard-Raten

Abbildung 22-14
Gesamttest für proportionale Hazard-Raten

Freiheitsgrade 1	Freiheitsgrade 2	Wald-F	Sig.
1	16	...	5.1358181E-5

Überlebenszeitvariable: Time to second arrest
Ereignisstatusvariable: Second arrest = 1
Modell: age, age*_TF

Abbildung 22-15
Parameterschätzer für alternatives Modell

Parameter	B	Standardfehler	90%-Konfidenzintervall	
			Untere Grenze	Obere Grenze
age	-,002	,014	-,025	,022
age*_TF[a]	-,012	,002	-,016	-,009

Überlebenszeitvariable: Time to second arrest
Ereignisstatusvariable: Second arrest = 1
Modell: age, age*_TF

a. Zeitfunktion: Log

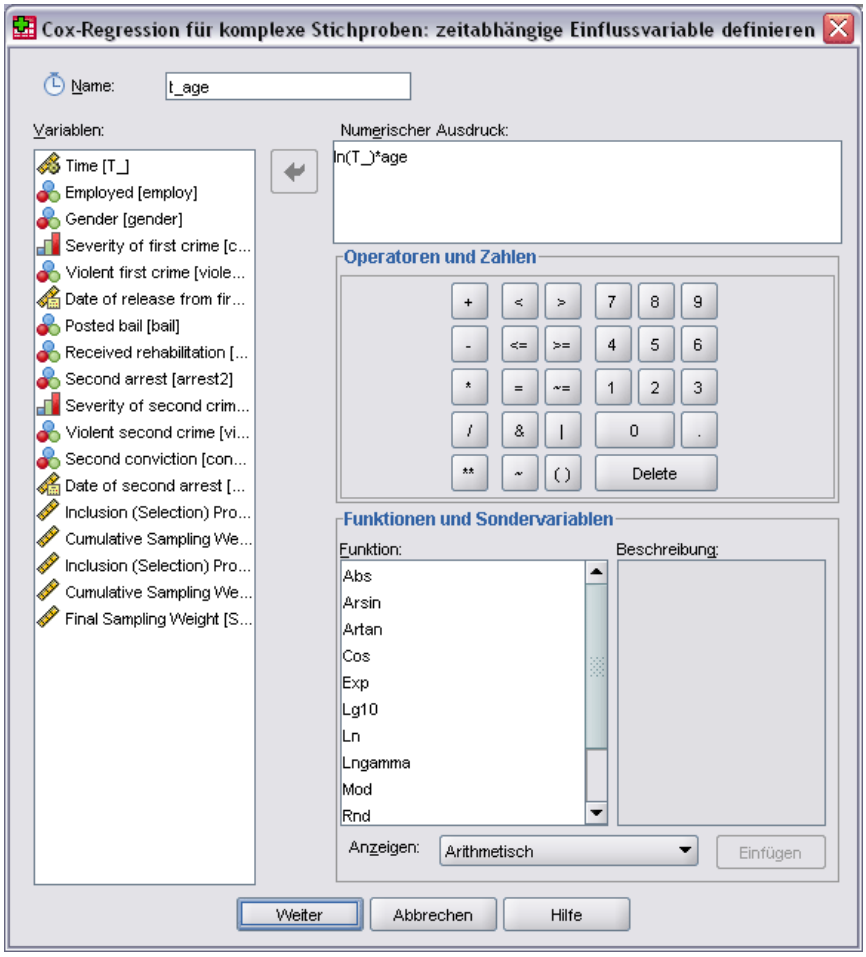
b. Entscheidungsmethode: Efron

Der Signifikanzwert des Gesamttests für proportionale Hazard-Raten liegt unter 0,05, was anzeigt, dass die Annahme proportionaler Hazard-Raten verletzt ist. Für das alternative Modell wird die Funktion zum Protokollieren der Zeit verwendet, wodurch die Reproduktion dieser zeitabhängigen Einflussvariablen erleichtert wird.

Hinzufügen einer zeitabhängigen Einflussvariablen

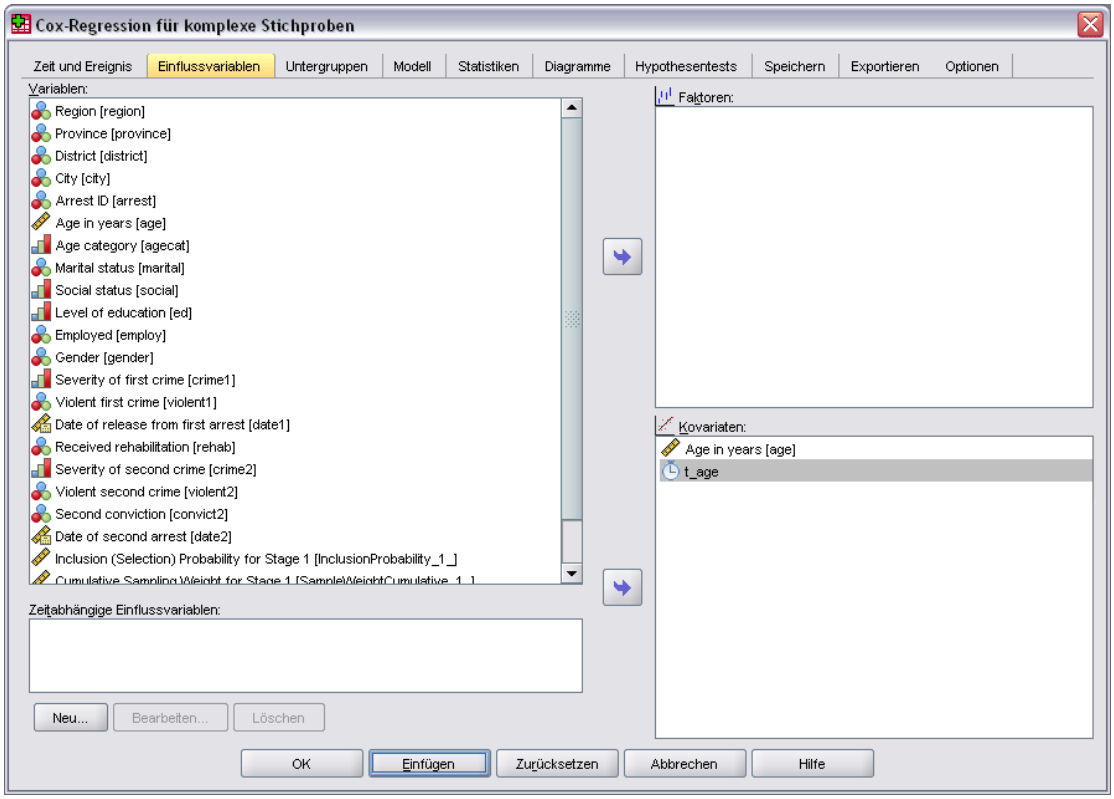
- Rufen Sie das Dialogfeld "Cox-Regression für komplexe Stichproben" erneut auf und klicken Sie auf die Registerkarte Einflussvariablen.
- Klicken Sie auf Neu.

Abbildung 22-16
Cox-Regression: Dialogfeld "Zeitabhängige Einflussvariable definieren"



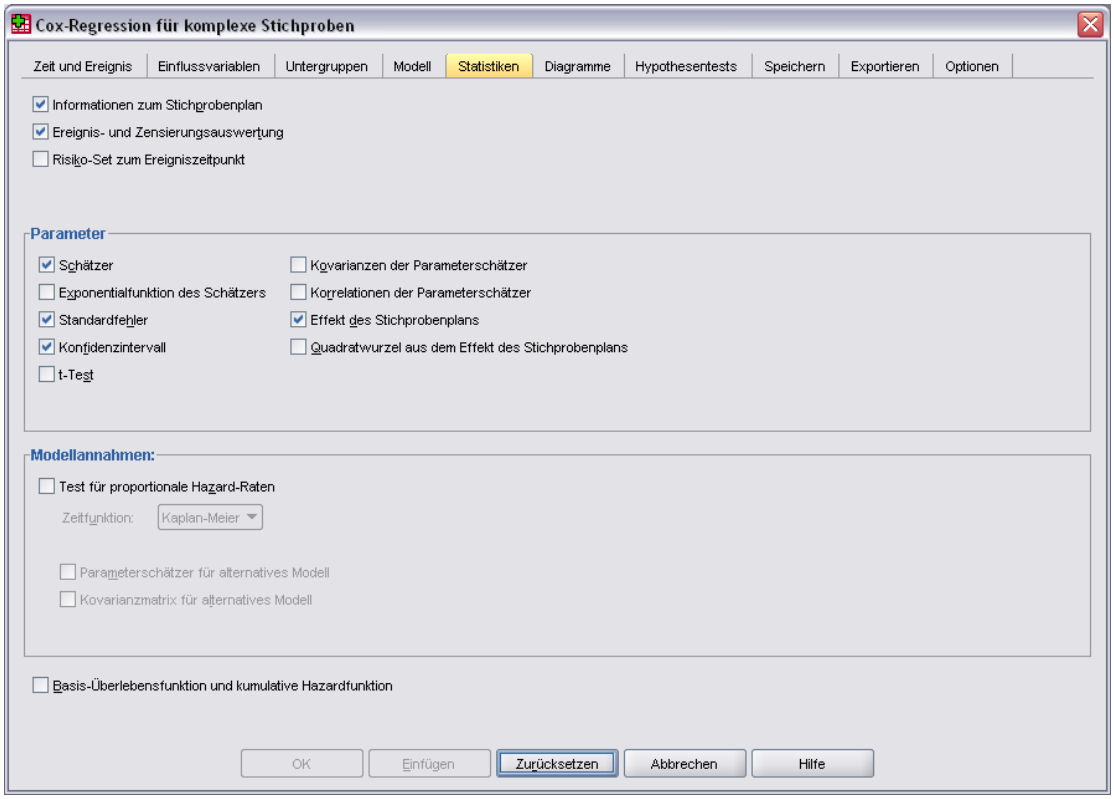
- Geben Sie t_age als Namen der zeitabhängigen Einflussvariablen ein, die definiert werden soll.
- Geben Sie ln(T_)*age als numerischen Ausdruck ein.
- Klicken Sie auf Weiter.

Abbildung 22-17
Dialogfeld "Cox-Regression", Registerkarte "Einflussvariablen"



- Wählen Sie *t_age* als Kovariate aus.
- Klicken Sie auf die Registerkarte Statistik.

Abbildung 22-18
Dialogfeld "Cox-Regression", Registerkarte "Statistiken"



- Wählen Sie Schätzer, Standardfehler, Konfidenzintervall und Effekt des Stichprobenplans in der Gruppe "Parameter" aus.
- Heben Sie im Gruppenfeld "Modellannahmen" die Auswahl von Test für proportionale Hazard-Raten und Parameterschätzer für alternatives Modell auf.
- Klicken Sie auf OK.

Tests der Modelleffekte

Abbildung 22-19
Tests der Modelleffekte

Quelle	Freiheitsgrade 1	Freiheitsgrade 2	Wald-F	Sig.
age	1,000	16,000	,015	0.91
t_age	1,000	16,000	29,924	1,580E-13

Überlebenszeitvariable: Time to second arrest
Ereignisstatusvariable: Second arrest = 1
Modell: age, t_age

Nach der Addition der zeitbasierten Einflussvariablen beträgt der Signifikanzwert für *age* (Alter) 0,91, was anzeigt, dass der Beitrag dieser Variablen zum Modell durch den Beitrag von *t_age* verdrängt wird.

Parameter-Schätzer

Abbildung 22-20
Parameterschätzer

Parameter	B	Standardfehler	90%-Konfidenzintervall		Effekt des Stichprobenplans
			Untere Grenze	Obere Grenze	
age	-,002	,001	-,030	,027	,702
age*_TF[a]	-,012	,002	-,017	-,008	,666

Überlebenszeitvariable: Time to second arrest
Ereignisstatusvariable: Second arrest = 1
Modell: age, age*_TF

Bei Betrachtung der Parameterschätzer und Standardfehler sehen Sie, dass Sie das alternative Modell aus dem Test für proportionale Hazard-Raten reproduziert haben. Durch die explizite Angabe des Modells können Sie weitere Parameterstatistiken und Diagramme anfordern. In diesem Fall haben wir den Effekt des Stichprobenplans angefordert; der Wert für *t_age* von weniger als 1 zeigt an, dass der Standardfehler für *t_age* kleiner ist als der Fehler, der sich unter der Annahme ergäbe, dass es sich bei dem Daten-Set um eine einfache Zufallsstichprobe handelt. In diesem Fall wäre der Effekt von *t_age* immer noch statistisch signifikant, die Konfidenzintervalle wären allerdings größer.

Mehrere Fälle pro Subjekt in "Cox-Regression für komplexe Stichproben"

Forscher befassen sich mit den Überlebenszeiten von Patienten, die nach einem Rehabilitationsprogramm wegen eines ischämischen Schlaganfalls mit einer Reihe von Problemen zu kämpfen haben.

Mehrere Fälle pro Subjekt. Variablen, die die Anamnese des Patienten repräsentieren, dürften als Einflussvariablen nützlich sein. Bei diesen Patienten werden im Laufe der Zeit möglicherweise wichtige medizinische Ereignisse eintreten, durch die ihre Anamnese verändert wird In diesem Daten-Set werden das Auftreten von Herzinfarkt, ischämischem Schlaganfall und hämorrhagischem Schlaganfall sowie der Zeitpunkt des Ereignisses aufgezeichnet. Sie könnten berechenbare, zeitabhängige Kovariaten innerhalb der Prozedur erstellen, um diese Informationen in das Modell aufzunehmen, es sollte jedoch praktischer sein, mehrere Fälle pro Subjekt zu verwenden. Beachten Sie, dass die Variablen ursprünglich kodiert waren, sodass die Aufzeichnungen über die Anamnese des Patienten auf verschiedene Variablen verteilt ist. Daher müssen Sie das Daten-Set umstrukturieren.

Abschneiden auf der linken Seite. Risikobeginn ist der Zeitpunkt des ischämischen Schlaganfalls. Die Stichprobe umfasst jedoch nur Patienten, die das Rehabilitationsprogramm überlebt haben; daher ist die Stichprobe auf der linken Seite in dem Sinne abgeschnitten, dass die beobachteten Überlebenszeiten durch die Rehabilitationslänge "erhöht" sind. Sie können dies berücksichtigen, indem Sie die Zeit angeben, zu der die Patienten die Rehabilitation zu Beginn der Studie beendet haben.

Kein Stichprobenplan. Das Daten-Set wurde nicht mithilfe eines komplexen Stichprobenplans erstellt und wird als einfache Zufallsstichprobe betrachtet. Um die Cox-Regression für komplexe Stichproben verwenden zu können, müssen Sie einen Analyseplan erstellen.

Das Daten-Set finden Sie in der Datei *stroke_survival.sav*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 271. Verwenden Sie den Assistenten für die Datenumstrukturierung, um die Daten für die Analyse vorzubereiten, und anschließend den Analysevorbereitungsassistenten, um einen einfachen Zufallsstichprobenplan zu erstellen, und schließlich die Cox-Regression für komplexe Stichproben, um ein Modell für die Überlebenszeiten zu erstellen.

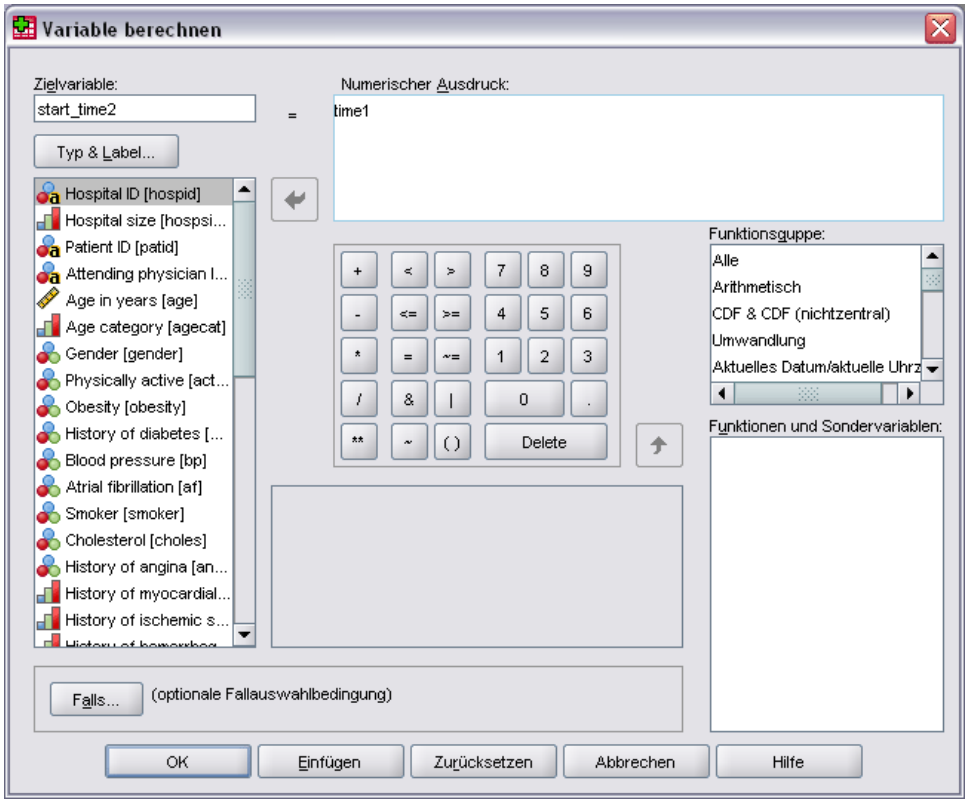
Vorbereiten der Daten für die Analyse

Vor der Umstrukturierung der Daten müssen Sie zwei Hilfsvariablen erstellen, die die Umstrukturierung unterstützen sollen.

▶ Wählen Sie zur Berechnung einer neuen Variablen folgende Optionen aus den Menüs aus:

Transformieren
Variable berechnen...

Abbildung 22-21
Dialogfeld "Variable berechnen"



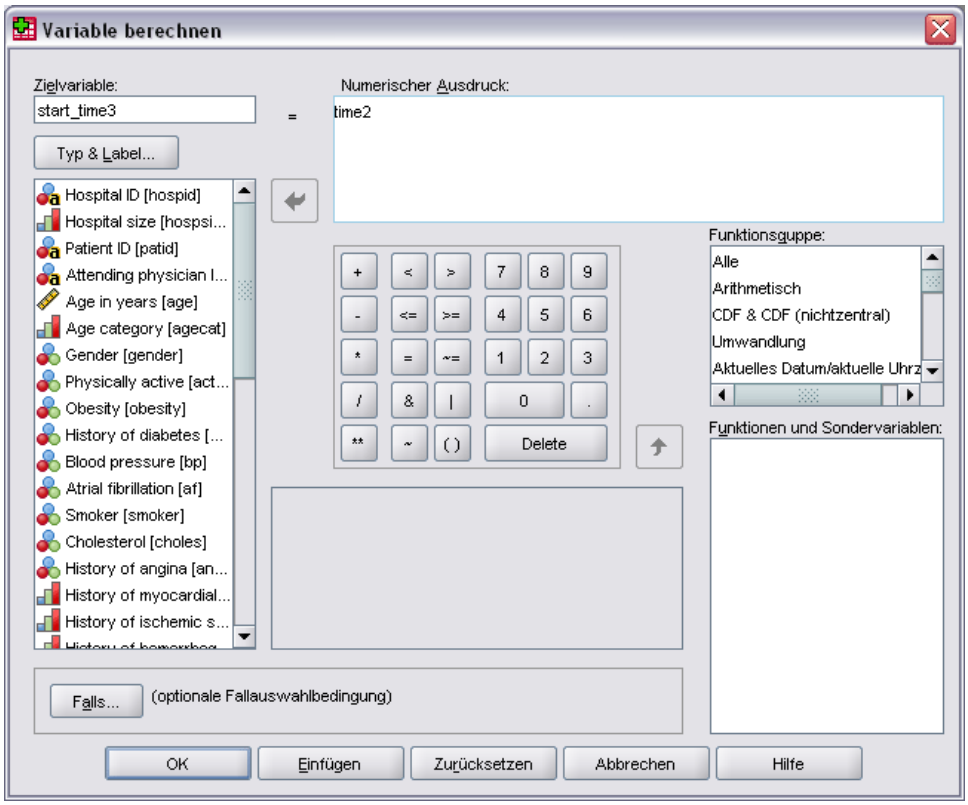
▶ Geben Sie *start_time2* als Zielvariable ein.

▶ Geben Sie *time1* als numerischen Ausdruck ein.

▶ Klicken Sie auf OK.

- Rufen Sie das Dialogfeld “Variable berechnen” auf.

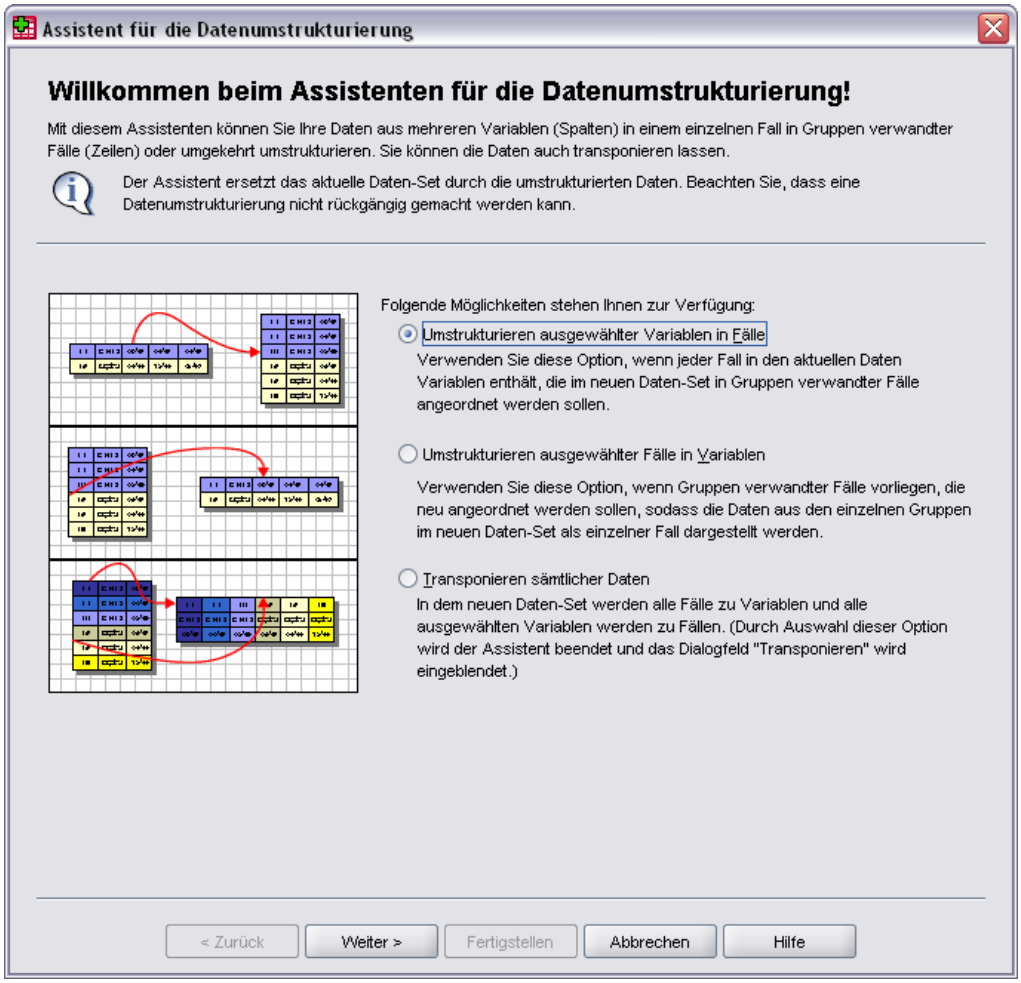
Abbildung 22-22
Dialogfeld “Variable berechnen”



- Geben Sie *start_time3* als Zielvariable ein.
- Geben Sie *time2* als numerischen Ausdruck ein.
- Klicken Sie auf OK.
- Zur Umstrukturierung der Daten von Variablen zu Fällen wählen Sie in den Menüs folgende Optionen aus:

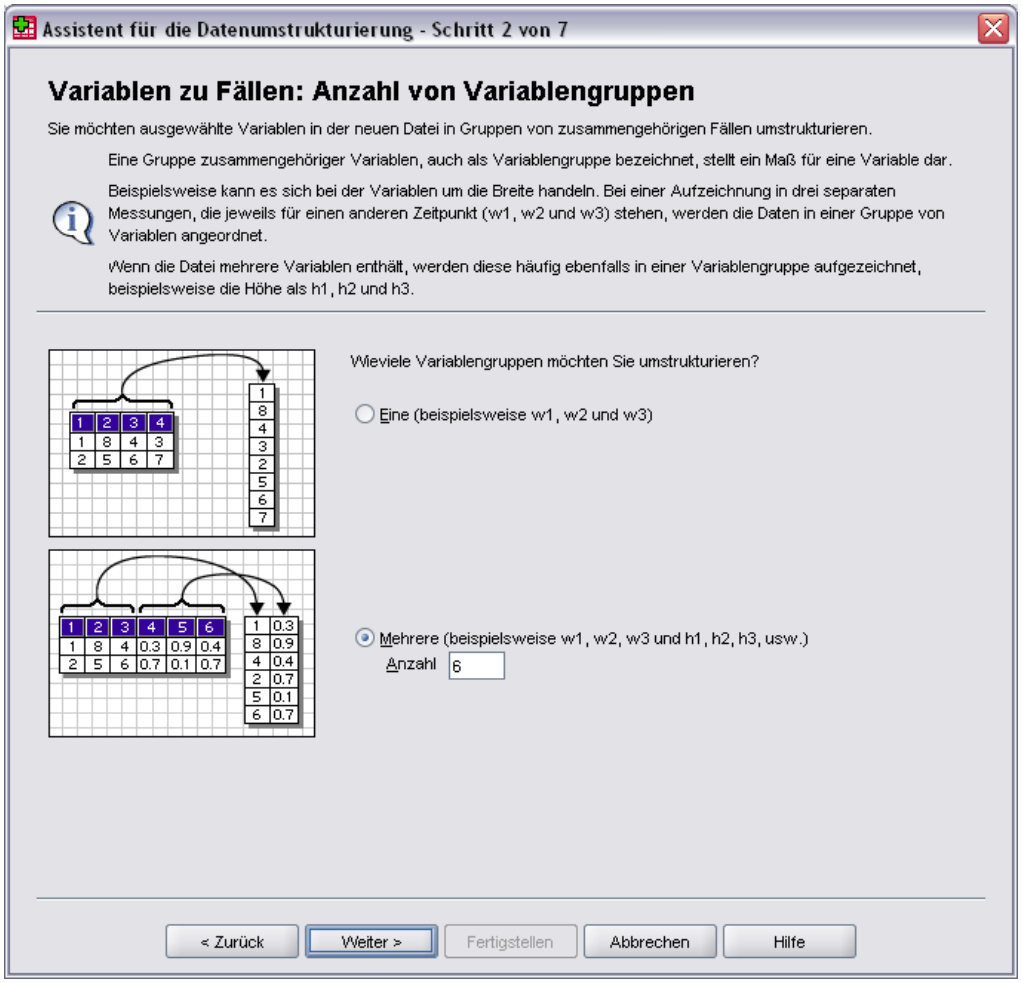
  Daten
  Umstrukturieren...

Abbildung 22-23
Assistent für die Datenumstrukturierung: Schritt "Willkommen"


Assistent für die Datenumstrukturierung
Willkommen beim Assistenten für die Datenumstrukturierung!
Mit diesem Assistenten können Sie Ihre Daten aus mehreren Variablen (Spalten) in einem einzelnen Fall in Gruppen verwandter Fälle (Zeilen) oder umgekehrt umstrukturieren. Sie können die Daten auch transponieren lassen.
Der Assistent ersetzt das aktuelle Daten-Set durch die umstrukturierten Daten. Beachten Sie, dass eine Datenumstrukturierung nicht rückgängig gemacht werden kann.
Folgende Möglichkeiten stehen Ihnen zur Verfügung:
Umstrukturieren ausgewählter Variablen in Fälle
Verwenden Sie diese Option, wenn jeder Fall in den aktuellen Daten Variablen enthält, die im neuen Daten-Set in Gruppen verwandter Fälle angeordnet werden sollen.
Umstrukturieren ausgewählter Fälle in Variablen
Verwenden Sie diese Option, wenn Gruppen verwandter Fälle vorliegen, die neu angeordnet werden sollen, sodass die Daten aus den einzelnen Gruppen im neuen Daten-Set als einzelner Fall dargestellt werden.
Transponieren sämtlicher Daten
In dem neuen Daten-Set werden alle Fälle zu Variablen und alle ausgewählten Variablen werden zu Fällen. (Durch Auswahl dieser Option wird der Assistent beendet und das Dialogfeld "Transponieren" wird eingeblendet.)
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

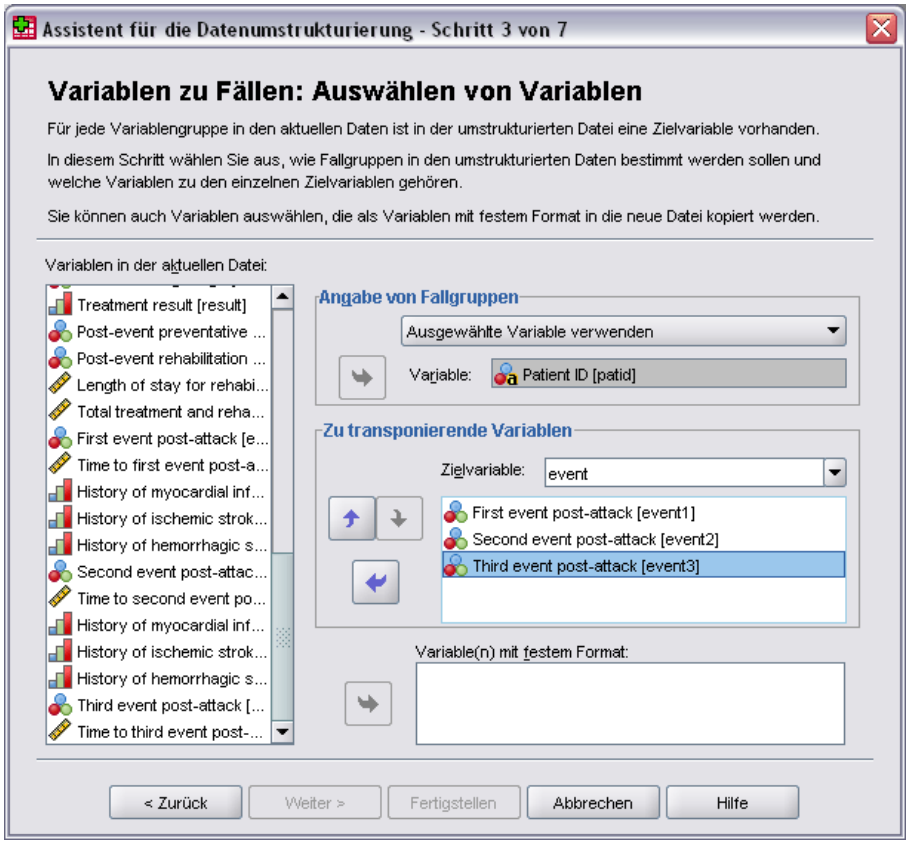
- Vergewissern Sie sich, dass Umstrukturieren ausgewählter Variablen in Fälle ausgewählt ist.
- Klicken Sie auf Weiter.

Abbildung 22-24
Assistent für die Datenumstrukturierung: Schritt "Variablen zu Fällen: Anzahl von Variablengruppen"


Assistent für die Datenumstrukturierung - Schritt 2 von 7
Variablen zu Fällen: Anzahl von Variablengruppen
Sie möchten ausgewählte Variablen in der neuen Datei in Gruppen von zusammengehörigen Fällen umstrukturieren.
Eine Gruppe zusammengehöriger Variablen, auch als Variablengruppe bezeichnet, stellt ein Maß für eine Variable dar.
Beispielsweise kann es sich bei der Variablen um die Breite handeln. Bei einer Aufzeichnung in drei separaten Messungen, die jeweils für einen anderen Zeitpunkt (w1, w2 und w3) stehen, werden die Daten in einer Gruppe von Variablen angeordnet.
Wenn die Datei mehrere Variablen enthält, werden diese häufig ebenfalls in einer Variablengruppe aufgezeichnet, beispielsweise die Höhe als h1, h2 und h3.
Wieviele Variablengruppen möchten Sie umstrukturieren?
Eine (beispielsweise w1, w2 und w3)
Mehrere (beispielsweise w1, w2, w3 und h1, h2, h3, usw.)
Anzahl 6
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

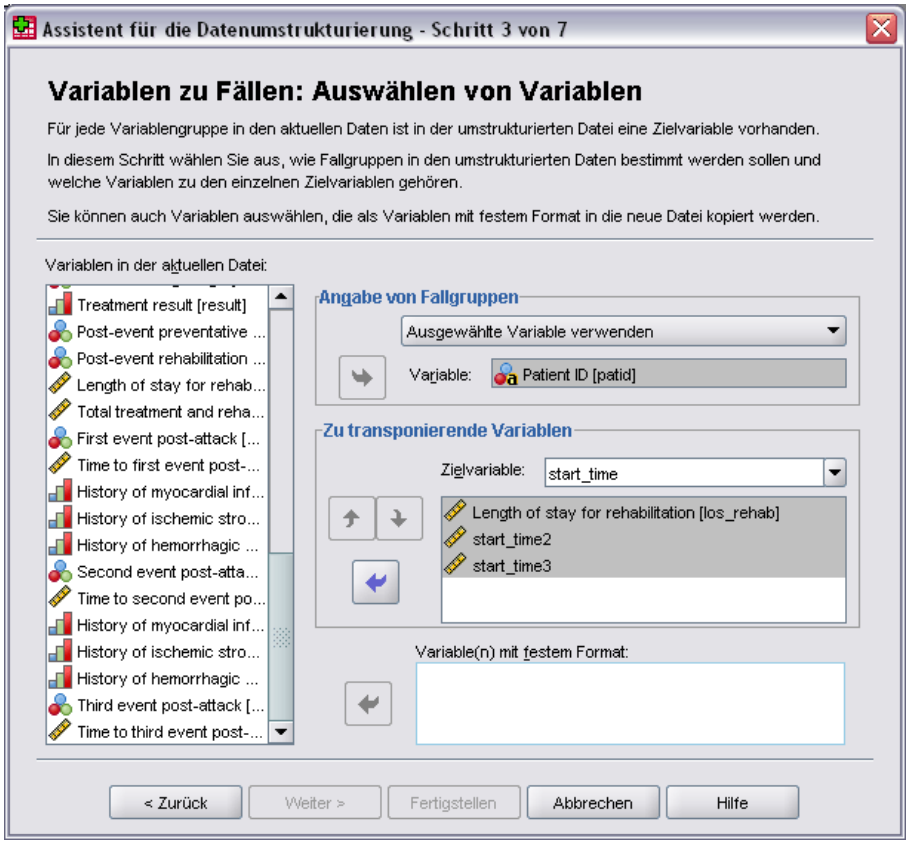
- Aktivieren Sie Mehrere für die Anzahl der umzustrukturierenden Variablengruppen.
- Geben Sie als Wert für die Anzahl der Gruppen 6 ein.
- Klicken Sie auf Weiter.

Abbildung 22-25
Assistent für die Datenumstrukturierung: Schritt "Variablen zu Fällen: Auswählen von Variablen"


Assistent für die Datenumstrukturierung - Schritt 3 von 7
Variablen zu Fällen: Auswählen von Variablen
Für jede Variablengruppe in den aktuellen Daten ist in der umstrukturierten Datei eine Zielvariable vorhanden.
In diesem Schritt wählen Sie aus, wie Fallgruppen in den umstrukturierten Daten bestimmt werden sollen und welche Variablen zu den einzelnen Zielvariablen gehören.
Sie können auch Variablen auswählen, die als Variablen mit festem Format in die neue Datei kopiert werden.
Variablen in der aktuellen Datei:
Treatment result [result]
Post-event preventative ...
Post-event rehabilitation ...
Length of stay for rehabi...
Total treatment and reha...
First event post-attack [e...
Time to first event post-a...
History of myocardial inf...
History of ischemic strok...
History of hemorrhagic s...
Second event post-attac...
Time to second event po...
History of myocardial inf...
History of ischemic strok...
History of hemorrhagic s...
Third event post-attack [...
Time to third event post-...
Angabe von Fallgruppen
Ausgewählte Variable verwenden
Variable:
Patient ID [patid]
Zu transponierende Variablen
Zielvariable:
event
First event post-attack [event1]
Second event post-attack [event2]
Third event post-attack [event3]
Variable(n) mit festem Format:
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

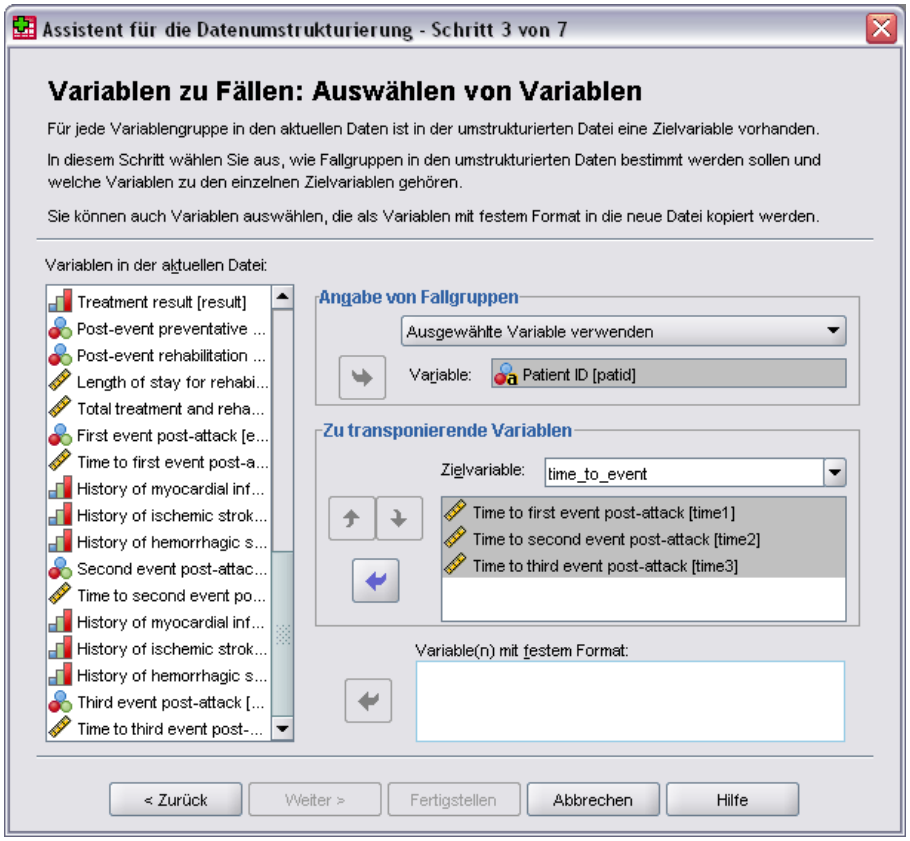
- Wählen Sie in der Gruppe "Angabe von Fallgruppen" Ausgewählte Variable verwenden und wählen Sie *Patient ID [patid]* (Patienten-ID) als Subjekt-ID.
- Geben Sie ereignis als erste Zielvariable ein.
- Wählen Sie *First event post-attack [event1]* (Erstes Ereignis nach dem Anfall), *Second event post-attack [event2]* (Zweites Ereignis nach dem Anfall) und *Third event post-attack [event3]* (Drittes Ereignis nach dem Anfall) als zu transponierende Variablen aus.
- Wählen Sie *trans2* aus der Liste der Zielvariablen.

Abbildung 22-26
Assistent für die Datenumstrukturierung: Schritt "Variablen zu Fällen: Auswählen von Variablen"


Assistent für die Datenumstrukturierung - Schritt 3 von 7
Variablen zu Fällen: Auswählen von Variablen
Für jede Variablengruppe in den aktuellen Daten ist in der umstrukturierten Datei eine Zielvariable vorhanden.
In diesem Schritt wählen Sie aus, wie Fallgruppen in den umstrukturierten Daten bestimmt werden sollen und welche Variablen zu den einzelnen Zielvariablen gehören.
Sie können auch Variablen auswählen, die als Variablen mit festem Format in die neue Datei kopiert werden.
Variablen in der aktuellen Datei:
Treatment result [result]
Post-event preventative ...
Post-event rehabilitation ...
Length of stay for rehab...
Total treatment and reha...
First event post-attack [...
Time to first event post-...
History of myocardial inf...
History of ischemic stro...
History of hemorrhagic ...
Second event post-atta...
Time to second event po...
History of myocardial inf...
History of ischemic stro...
History of hemorrhagic ...
Third event post-attack [...
Time to third event post-...
Angabe von Fallgruppen
Ausgewählte Variable verwenden
Variable:
Patient ID [patid]
Zu transponierende Variablen
Zielvariable:
start_time
Length of stay for rehabilitation [los_rehab]
start_time2
start_time3
Variable(n) mit festem Format:
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

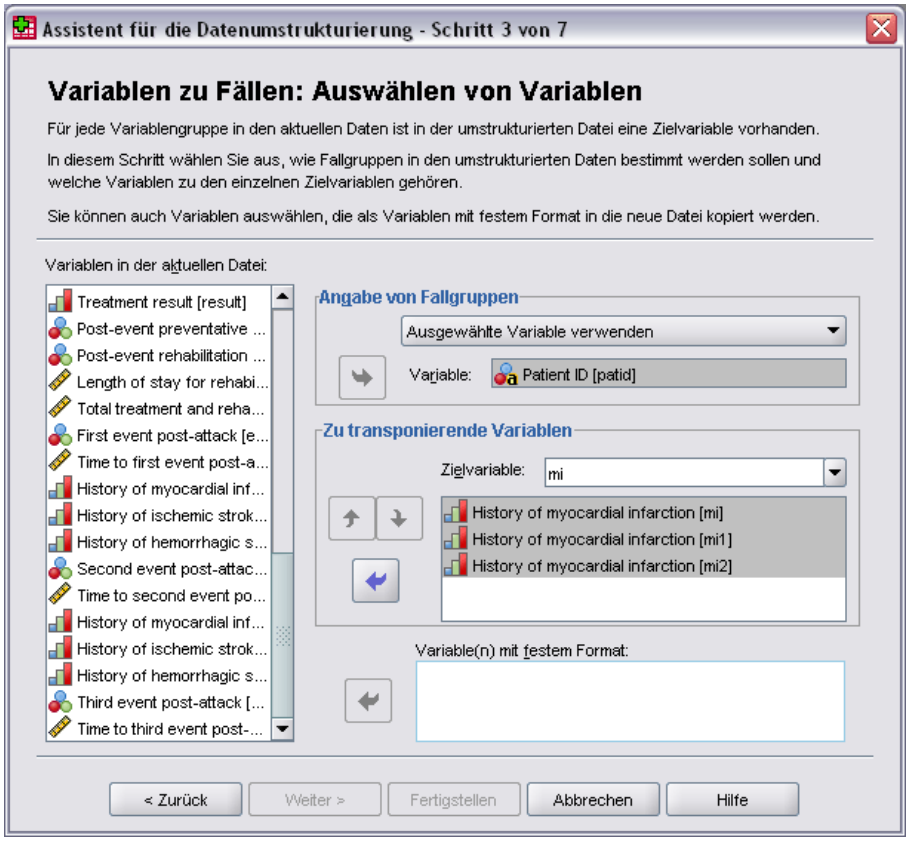
- Geben Sie start_time als Zielvariable ein.
- Wählen Sie *Length of stay for rehabilitation [los_rehab]* (Aufenthaltsdauer für Rehabilitation), *start_time2* und *start_time3* als zu transponierende Variablen aus. *Time to first event post-attack [time1]* (Zeit bis zum ersten Ereignis nach dem Anfall) und *Time to second event post-attack [time2]* (Zeit bis zum zweiten Ereignis nach dem Anfall) werden zum Erstellen der Endzeitpunkte verwendet und jede Variable kann nur in einer Liste von zu transponierenden Variablen vorkommen. Daher waren *start_time2* und *start_time3* erforderlich.
- Wählen Sie *trans3* aus der Liste der Zielvariablen.

Abbildung 22-27
Assistent für die Datenumstrukturierung: Schritt "Variablen zu Fällen: Auswählen von Variablen"


Assistent für die Datenumstrukturierung - Schritt 3 von 7
Variablen zu Fällen: Auswählen von Variablen
Für jede Variablengruppe in den aktuellen Daten ist in der umstrukturierten Datei eine Zielvariable vorhanden.
In diesem Schritt wählen Sie aus, wie Fallgruppen in den umstrukturierten Daten bestimmt werden sollen und welche Variablen zu den einzelnen Zielvariablen gehören.
Sie können auch Variablen auswählen, die als Variablen mit festem Format in die neue Datei kopiert werden.
Variablen in der aktuellen Datei:
Treatment result [result]
Post-event preventative ...
Post-event rehabilitation ...
Length of stay for rehabi...
Total treatment and reha...
First event post-attack [e...
Time to first event post-a...
History of myocardial inf...
History of ischemic strok...
History of hemorrhagic s...
Second event post-attac...
Time to second event po...
History of myocardial inf...
History of ischemic strok...
History of hemorrhagic s...
Third event post-attack [...
Time to third event post-...
Angabe von Fallgruppen
Ausgewählte Variable verwenden
Variable:
Patient ID [patid]
Zu transponierende Variablen
Zielvariable:
time_to_event
Time to first event post-attack [time1]
Time to second event post-attack [time2]
Time to third event post-attack [time3]
Variable(n) mit festem Format:
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

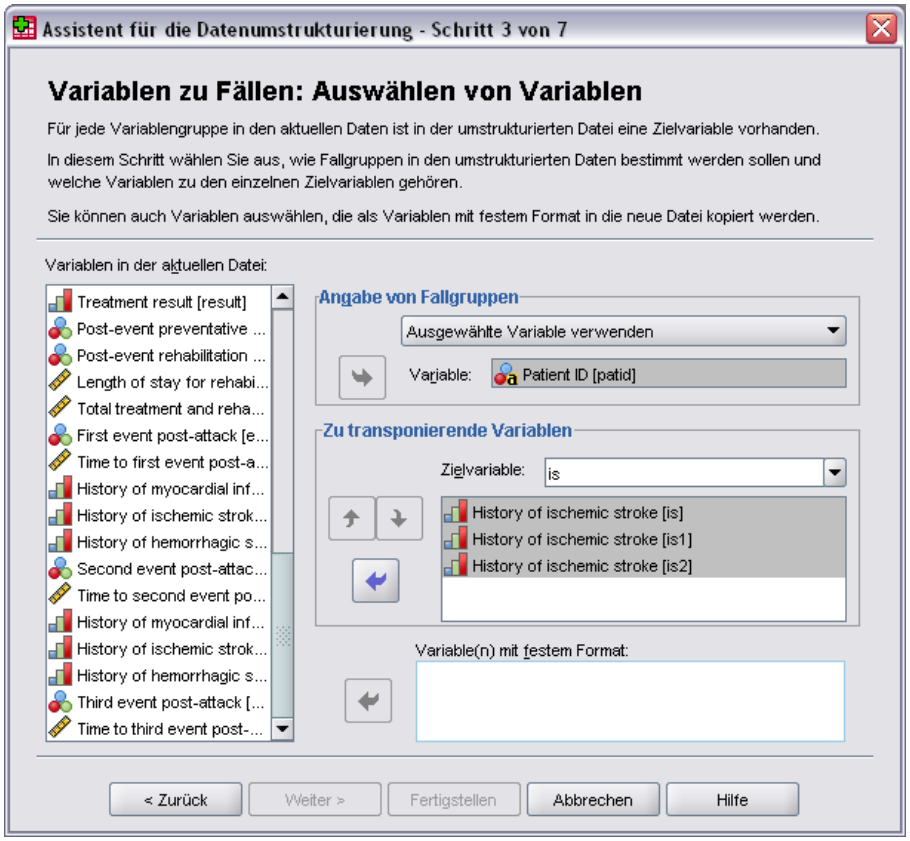
- Geben Sie time_to_event als Zielvariable ein.
- Wählen Sie *First event post-attack [event1]* (Erstes Ereignis nach dem Anfall), *Second event post-attack [event2]* (Zweites Ereignis nach dem Anfall) und *Third event post-attack [event3]* (Drittes Ereignis nach dem Anfall) als zu transponierende Variablen aus.
- Wählen Sie *trans4* aus der Liste der Zielvariablen.

Abbildung 22-28
Assistent für die Datenumstrukturierung: Schritt "Variablen zu Fällen: Auswählen von Variablen"


Assistent für die Datenumstrukturierung - Schritt 3 von 7
Variablen zu Fällen: Auswählen von Variablen
Für jede Variablengruppe in den aktuellen Daten ist in der umstrukturierten Datei eine Zielvariable vorhanden.
In diesem Schritt wählen Sie aus, wie Fallgruppen in den umstrukturierten Daten bestimmt werden sollen und welche Variablen zu den einzelnen Zielvariablen gehören.
Sie können auch Variablen auswählen, die als Variablen mit festem Format in die neue Datei kopiert werden.
Variablen in der aktuellen Datei:
Treatment result [result]
Post-event preventative ...
Post-event rehabilitation ...
Length of stay for rehabi...
Total treatment and reha...
First event post-attack [e...
Time to first event post-a...
History of myocardial inf...
History of ischemic strok...
History of hemorrhagic s...
Second event post-attac...
Time to second event po...
History of myocardial inf...
History of ischemic strok...
History of hemorrhagic s...
Third event post-attack [...
Time to third event post-...
Angabe von Fallgruppen
Ausgewählte Variable verwenden
Variable:
Patient ID [patid]
Zu transponierende Variablen
Zielvariable:
mi
History of myocardial infarction [mi]
History of myocardial infarction [mi1]
History of myocardial infarction [mi2]
Variable(n) mit festem Format:
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

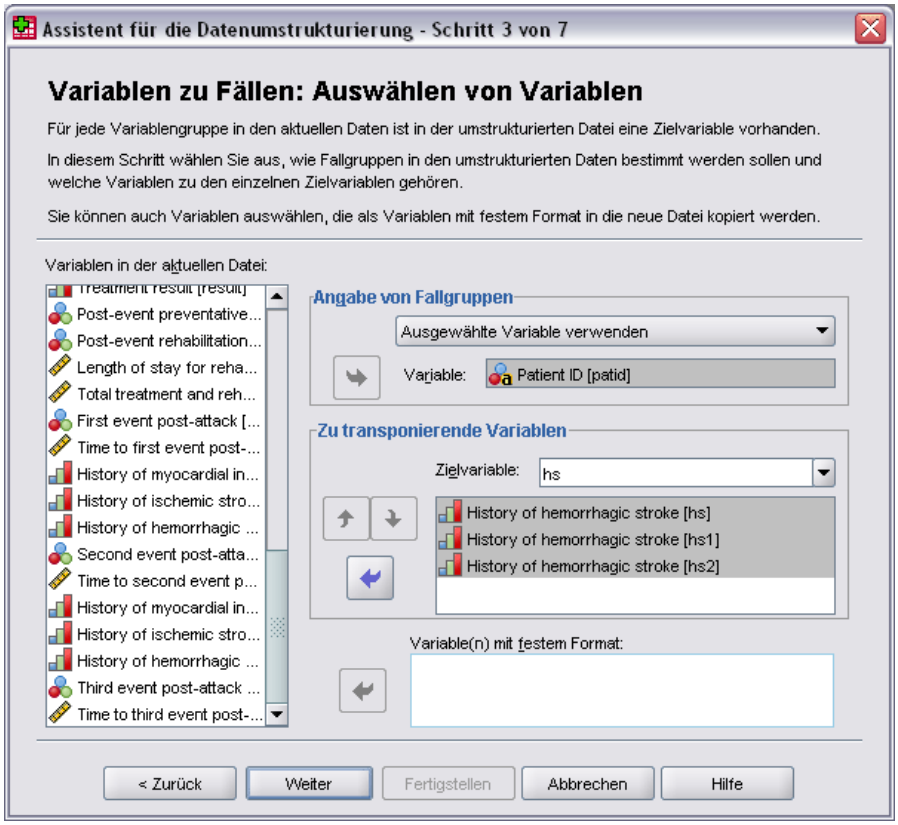
- Geben Sie mi als Zielvariable ein.
- Wählen Sie *History of myocardial infarction [mi]* (Herzinfarktanamnese), *History of myocardial infarction [mi1]* und *History of myocardial infarction [mi2]* als zu transponierende Variablen aus.
- Wählen Sie *trans5* aus der Liste der Zielvariablen.

Abbildung 22-29
Assistent für die Datenumstrukturierung: Schritt "Variablen zu Fällen: Auswählen von Variablen"


Assistent für die Datenumstrukturierung - Schritt 3 von 7
Variablen zu Fällen: Auswählen von Variablen
Für jede Variablengruppe in den aktuellen Daten ist in der umstrukturierten Datei eine Zielvariable vorhanden.
In diesem Schritt wählen Sie aus, wie Fallgruppen in den umstrukturierten Daten bestimmt werden sollen und welche Variablen zu den einzelnen Zielvariablen gehören.
Sie können auch Variablen auswählen, die als Variablen mit festem Format in die neue Datei kopiert werden.
Variablen in der aktuellen Datei:
Treatment result [result]
Post-event preventative ...
Post-event rehabilitation ...
Length of stay for rehabi...
Total treatment and reha...
First event post-attack [e...
Time to first event post-a...
History of myocardial inf...
History of ischemic strok...
History of hemorrhagic s...
Second event post-attac...
Time to second event po...
History of myocardial inf...
History of ischemic strok...
History of hemorrhagic s...
Third event post-attack [...
Time to third event post-...
Angabe von Fallgruppen
Ausgewählte Variable verwenden
Variable:
Patient ID [patid]
Zu transponierende Variablen
Zielvariable:
is
History of ischemic stroke [is]
History of ischemic stroke [is1]
History of ischemic stroke [is2]
Variable(n) mit festem Format:
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

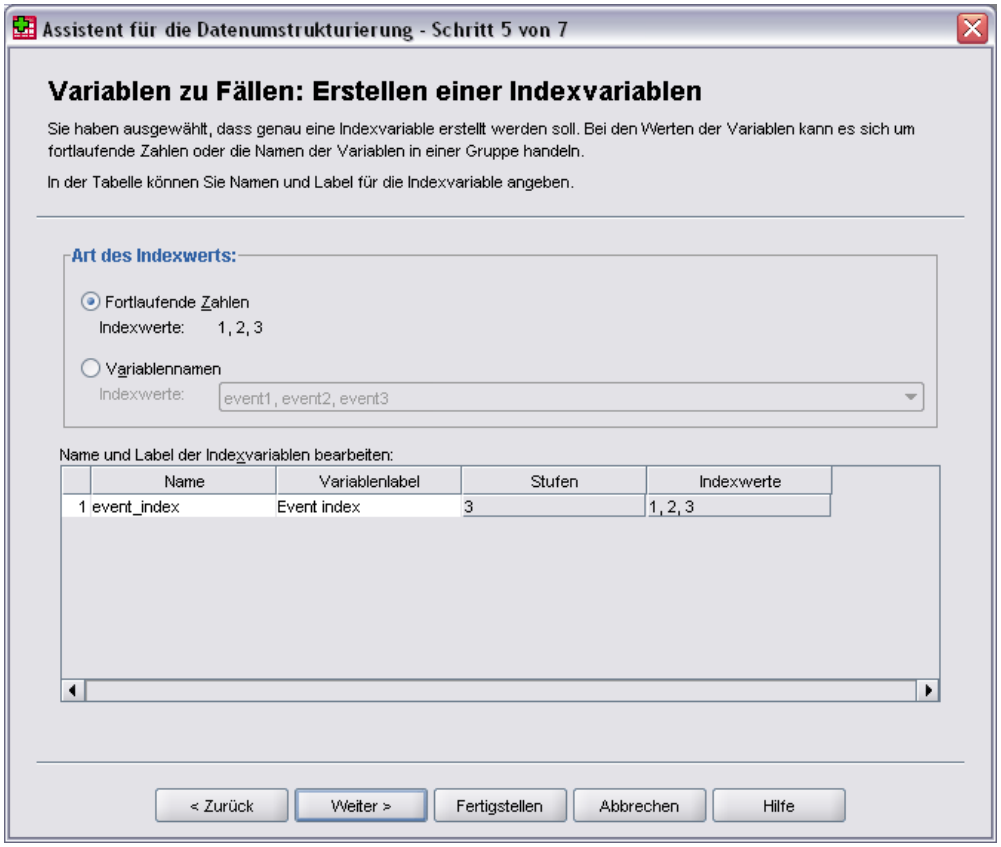
- Geben Sie is als Zielvariable ein.
- Wählen Sie *History of ischemic stroke [is]* (ischämische Schlaganfälle in der Anamnese), *History of ischemic stroke [is1]* und *History of ischemic stroke [is2]* als zu transponierende Variablen auf
- Wählen Sie *trans6* aus der Liste der Zielvariablen.

Abbildung 22-30
Assistent für die Datenumstrukturierung: Schritt "Variablen zu Fällen: Auswählen von Variablen"


Assistent für die Datenumstrukturierung - Schritt 3 von 7
Variablen zu Fällen: Auswählen von Variablen
Für jede Variablengruppe in den aktuellen Daten ist in der umstrukturierten Datei eine Zielvariable vorhanden.
In diesem Schritt wählen Sie aus, wie Fallgruppen in den umstrukturierten Daten bestimmt werden sollen und welche Variablen zu den einzelnen Zielvariablen gehören.
Sie können auch Variablen auswählen, die als Variablen mit festem Format in die neue Datei kopiert werden.
Variablen in der aktuellen Datei:
Post-event preventative...
Post-event rehabilitation...
Length of stay for reha...
Total treatment and reh...
First event post-attack [...
Time to first event post-...
History of myocardial in...
History of ischemic stro...
History of hemorrhagic ...
Second event post-atta...
Time to second event p...
History of myocardial in...
History of ischemic stro...
History of hemorrhagic ...
Third event post-attack ...
Time to third event post-...
Angabe von Fallgruppen
Ausgewählte Variable verwenden
Variable:
Patient ID [patid]
Zu transponierende Variablen
Zielvariable:
hs
History of hemorrhagic stroke [hs]
History of hemorrhagic stroke [hs1]
History of hemorrhagic stroke [hs2]
Variable(n) mit festem Format:
< Zurück
Weiter
Fertigstellen
Abbrechen
Hilfe

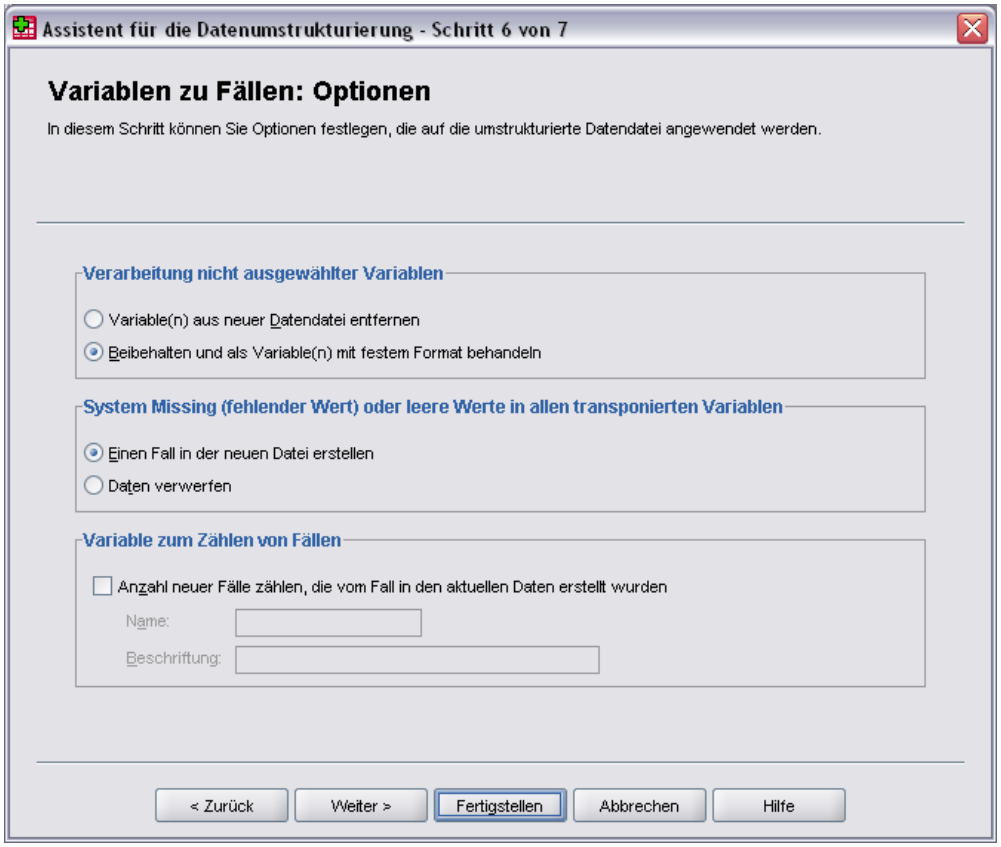
- Geben Sie hs als Zielvariable ein.
- Wählen Sie *History of hemorrhagic stroke [hs]* (hämorrhagische Schlaganfälle in der Anamnese), *History of hemorrhagic stroke [hs1]* und *History of hemorrhagic stroke [hs2]* als zu transponierende Variablen aus.
- Klicken Sie auf Weiter und klicken Sie anschließend im Schritt "Erstellen von Indexvariablen" ebenfalls auf Weiter.

Abbildung 22-31
Assistent für die Datenumstrukturierung: Schritt "Variablen zu Fällen: Optionen"


Assistent für die Datenumstrukturierung - Schritt 5 von 7
Variablen zu Fällen: Erstellen einer Indexvariablen
Sie haben ausgewählt, dass genau eine Indexvariable erstellt werden soll. Bei den Werten der Variablen kann es sich um fortlaufende Zahlen oder die Namen der Variablen in einer Gruppe handeln.
In der Tabelle können Sie Namen und Label für die Indexvariable angeben.
Art des Indexwerts:
Fortlaufende Zahlen
Indexwerte: 1, 2, 3
Variablennamen
Indexwerte: event1, event2, event3
Name und Label der Indexvariablen bearbeiten:
Name
Variablenlabel
Stufen
Indexwerte
1 event_index
Event index
3
1, 2, 3
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

- Geben Sie event_index als Namen der Indexvariablen und Event index (Ereignisindex) als Variablenlabel ein.
- Klicken Sie auf Weiter.

Abbildung 22-32
Assistent für die Datenumstrukturierung: Schritt "Variablen zu Fällen: Optionen"


Assistent für die Datenumstrukturierung - Schritt 6 von 7
Variablen zu Fällen: Optionen
In diesem Schritt können Sie Optionen festlegen, die auf die umstrukturierte Datendatei angewendet werden.
Verarbeitung nicht ausgewählter Variablen
Variable(n) aus neuer Datendatei entfernen
Beibehalten und als Variable(n) mit festem Format behandeln
System Missing (fehlender Wert) oder leere Werte in allen transponierten Variablen
Einen Fall in der neuen Datei erstellen
Daten verwerfen
Variable zum Zählen von Fällen
Anzahl neuer Fälle zählen, die vom Fall in den aktuellen Daten erstellt wurden
Name:
Beschriftung:
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

- Vergewissern Sie sich, dass Beibehalten und als Variable(n) mit festem Format behandeln ausgewählt ist.
- Klicken Sie auf Fertig stellen.

Abbildung 22-33
Umstrukturieren von Daten

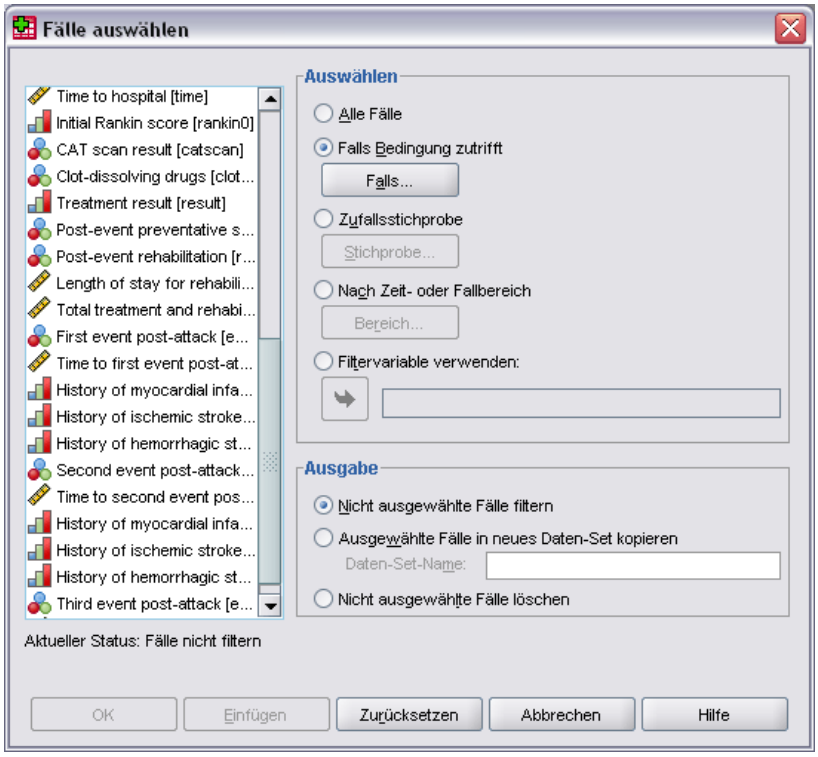
event_index	event	start_time	time_to_event	mi	is	hs
1	0	3	1500	0	1	0
2	-4	1500	-4	-4	-4	-4
3	-4	.	-4	-4	-4	-4
1	1	33	1311	0	1	0
2	4	1311	1325	1	1	0
3	-3	1325	-3	-3	-3	-3
1	4	12	1098	1	1	0
2	-3	1098	-3	-3	-3	-3
3	-3	.	-3	-3	-3	-3
1	4	4	1356	0	1	0
2	-3	1356	-3	-3	-3	-3
3	-3	.	-3	-3	-3	-3

Die umstrukturierten Daten enthalten drei Fälle für jeden Patienten; bei vielen Patienten traten jedoch weniger als drei Ereignisse auf, sodass es viele Fälle mit negativen (fehlenden) Werten für *event* gibt. Sie können diese Fälle einfach aus dem Daten-Set herausfiltern.

► Um diese Fälle herauszufiltern, wählen Sie die folgenden Optionen aus den Menüs aus:

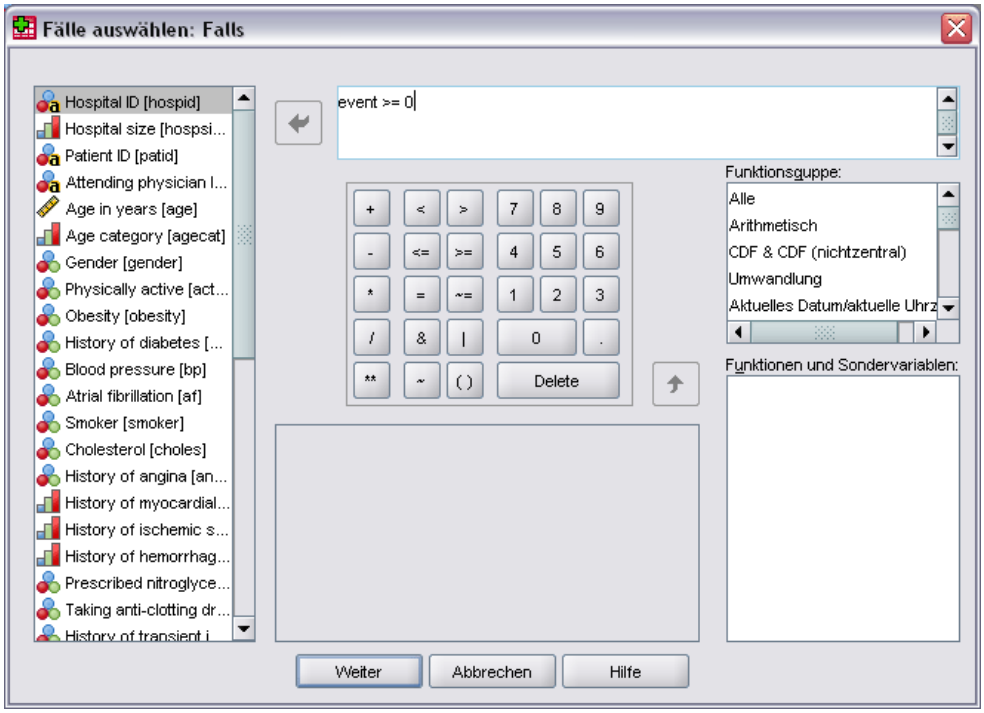
Daten
Fälle auswählen...

Abbildung 22-34
Dialogfeld "Fälle auswählen"



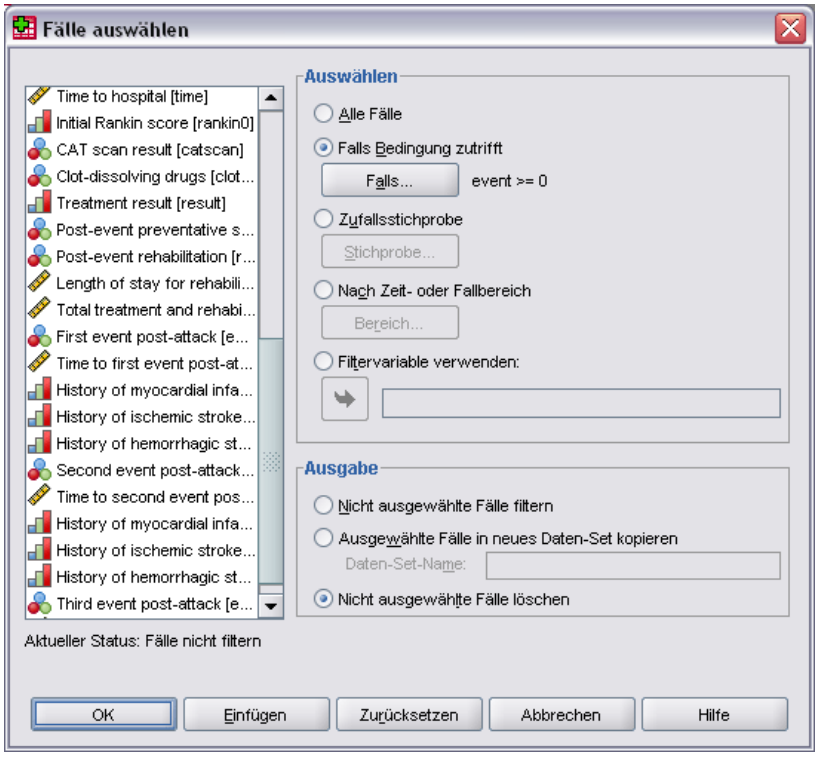
- Aktivieren Sie das Optionsfeld Falls Bedingung zutrifft.
- Klicken Sie auf Falls.

Abbildung 22-35
Dialogfeld "Fälle auswählen: Falls"



- Geben Sie event >= 0 als Ausdruck für die Bedingung ein.
- Klicken Sie auf Weiter.

Abbildung 22-36
Dialogfeld "Fälle auswählen"



- Wählen Sie Nicht ausgewählte Fälle löschen.
- Klicken Sie auf OK.

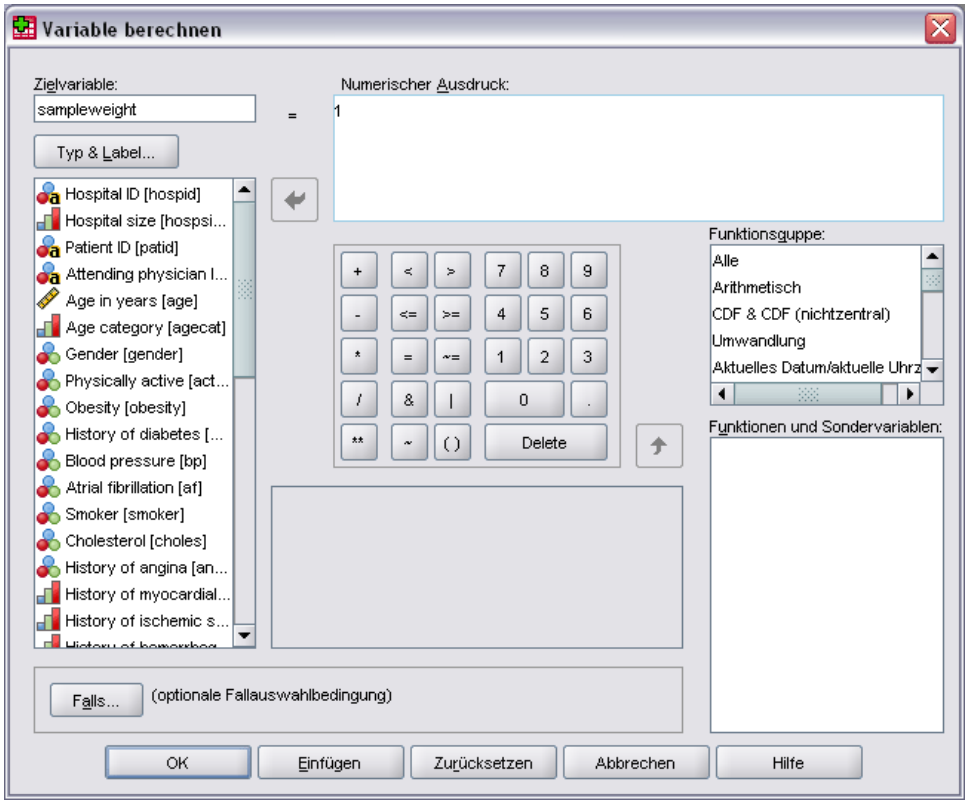
Erstellen eines Analyseplans für einfache Zufallsstichprobenziehungen

Nun können Sie den Analyseplan für einfache Zufallsstichprobenziehungen erstellen.

- Zunächst müssen Sie eine Variable für die Stichprobengewichtung erstellen. Wählen Sie die folgenden Befehle aus den Menüs aus:

  Transformieren
  Variable berechnen...

Abbildung 22-37
Dialogfeld "Variable berechnen"



- Geben Sie sampleweight als Zielvariable ein.
- Geben Sie 1 als numerischen Ausdruck ein.
- Klicken Sie auf OK.

Nun können Sie den Analyseplan erstellen.

Anmerkung: Im Verzeichnis mit den Beispieldateien befindet sich eine bestehende Plandatei, *srs.csaplan*, die Sie verwenden können, wenn Sie die folgenden Anweisungen überspringen und gleich mit der Datenanalyse fortfahren möchten.

- Zum Erstellen des Analyseplans wählen Sie folgende Optionen in den Menüs aus:

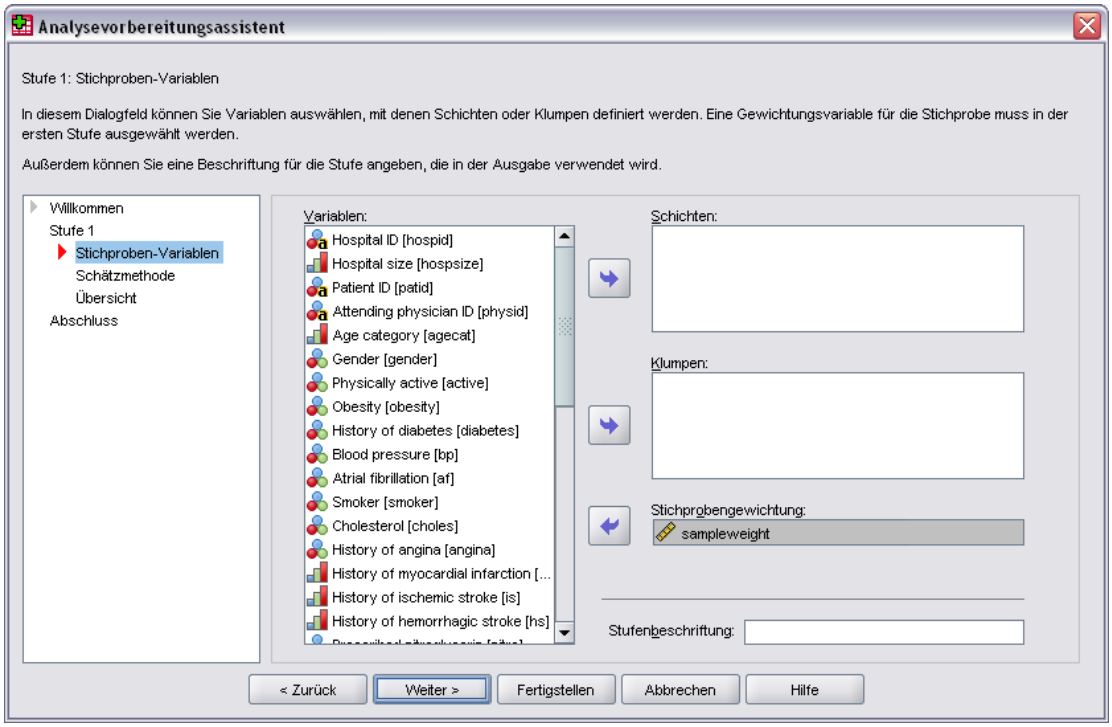
Analysieren
  Komplexe Stichproben
    Für Analyse vorbereiten...

Abbildung 22-38
Analysevorbereitungsassistent – Schritt "Willkommen"

Analysevorbereitungsassistent
Willkommen beim Analysevorbereitungsassistenten
Der Analysevorbereitungsassistent unterstützt Sie bei der Beschreibung Ihrer komplexen Stichprobe und bei der Auswahl einer Schätzmethode. Sie werden aufgefordert, Stichprobengewichtungen und andere Informationen bereitzustellen, die für eine genaue Schätzung der Standardfehler erforderlich sind.
Die getroffene Auswahl wird in einer Plandatei gespeichert, die Sie bei allen Analyseverfahren der Option "Komplexe Stichproben" verwenden können.
Wie möchten Sie vorgehen?
Plandatei erstellen
Wählen Sie diese Option, wenn Sie über Stichprobendaten verfügen, jedoch noch keine Plandatei erstellt haben.
Datei:
srs.csaplan
Durchsuchen...
Plandatei bearbeiten
Wählen Sie diese Option, wenn Sie Stufen eines bestehenden Plans hinzufügen, entfernen oder bearbeiten möchten.
Datei
Durchsuchen...
Wenn Sie bereits über eine Plandatei verfügen, können Sie den Analysevorbereitungsassistenten übergehen und direkt mit einer der Analyseprozeduren in der Option "Komplexe Stichproben" fortfahren, um Ihre Stichprobe zu analysieren.
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

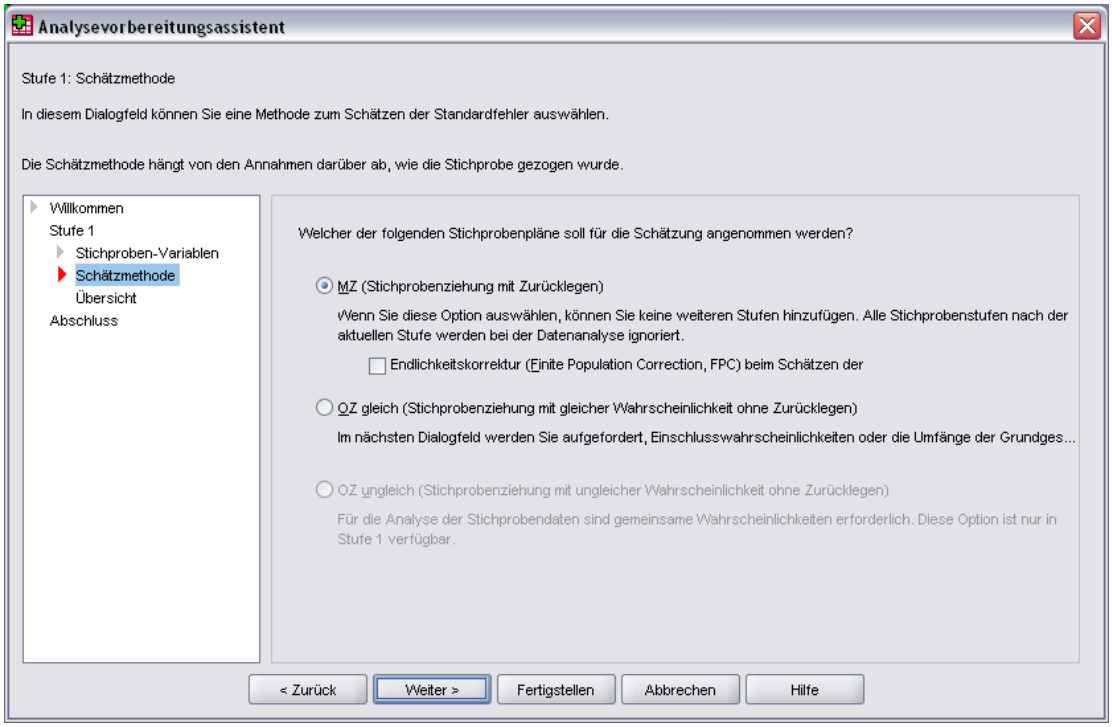
- Wählen Sie die Option Plandatei erstellen und geben Sie als Namen der Datei srs.csaplan ein. Wechseln Sie gegebenenfalls zu dem gewünschten Speicherort.
- Klicken Sie auf Weiter.

Abbildung 22-39
Analysevorbereitungsassistent – Stichproben-Variablen



- Wählen Sie *sampleweight* als Variable für die Stichprobengewichtung aus.
- Klicken Sie auf Weiter.

Abbildung 22-40
Analysevorbereitungsassistent – Schätzmethode


Analysevorbereitungsassistent
Stufe 1: Schätzmethode
In diesem Dialogfeld können Sie eine Methode zum Schätzen der Standardfehler auswählen.
Die Schätzmethode hängt von den Annahmen darüber ab, wie die Stichprobe gezogen wurde.
Willkommen
Stufe 1
Stichproben-Variablen
Schätzmethode
Übersicht
Abschluss
Welcher der folgenden Stichprobenpläne soll für die Schätzung angenommen werden?
MZ (Stichprobenziehung mit Zurücklegen)
Wenn Sie diese Option auswählen, können Sie keine weiteren Stufen hinzufügen. Alle Stichprobenstufen nach der aktuellen Stufe werden bei der Datenanalyse ignoriert.
Endlichkeitskorrektur (Finite Population Correction, FPC) beim Schätzen der
OZ gleich (Stichprobenziehung mit gleicher Wahrscheinlichkeit ohne Zurücklegen)
Im nächsten Dialogfeld werden Sie aufgefordert, Einschlusswahrscheinlichkeiten oder die Umfänge der Grundges...
OZ ungleich (Stichprobenziehung mit ungleicher Wahrscheinlichkeit ohne Zurücklegen)
Für die Analyse der Stichprobendaten sind gemeinsame Wahrscheinlichkeiten erforderlich. Diese Option ist nur in Stufe 1 verfügbar.
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

- Deaktivieren Sie die Option Endlichkeitskorrektur [...] verwenden.
- Klicken Sie auf Fertig stellen.

Nun können Sie die Analyse durchführen.

Durchführung der Analyse

- Um eine Analyse der Art “Cox-Regression für komplexe Stichproben” durchzuführen, wählen Sie folgende Optionen aus den Menüs aus:

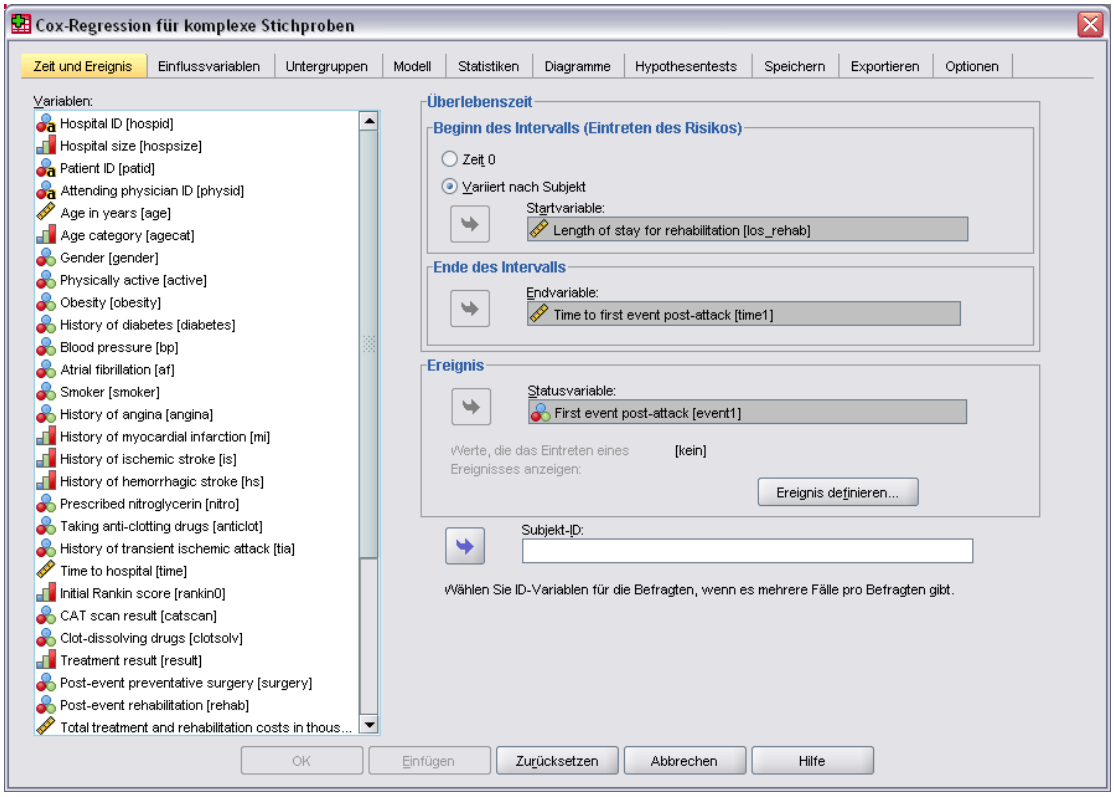
Analysieren
  Komplexe Stichproben
    Cox-Regression…

Abbildung 22-41
Dialogfeld "Plan für Cox-Regression"

Plan für komplexe Stichproben für Cox-Regression
Plan
Datei:
srs.csaplan
Durchsuchen
Wenn Sie nicht über eine Plandatei für Ihre komplexe Stichprobe verfügen, können Sie eine Plandatei mithilfe des Analysevorbereitungsassistenten erstellen. Wählen Sie im Menü "Komplexe Stichproben" die Option "Für Analyse vorbereiten", um den Assistenten aufzurufen.
Gemeinsame Wahrscheinlichkeiten
Gemeinsame Wahrscheinlichkeiten sind erforderlich, wenn für den Plan eine Schätzung für ungleiche Wahrscheinlichkeiten OZ erforderlich ist. Andernfalls werden sie ignoriert.
Standarddatei verwenden (srs.sav)
Ein offenes Daten-Set
stroke_survival.sav []
Benutzerdefinierte Datei
Datei:
Durchsuchen...
Weiter
Abbrechen
Hilfe

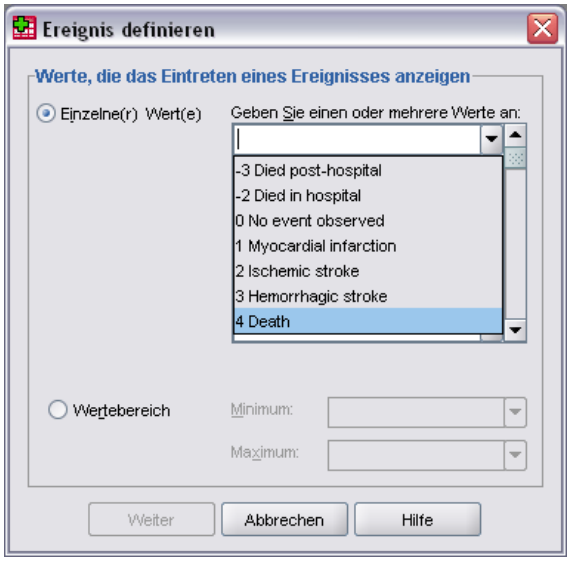
- Wechseln Sie in das Verzeichnis, in dem Sie den Analyseplan für einfache Zufallsstichprobenziehungen gespeichert haben, bzw. zum Verzeichnis mit den Beispieldateien und wählen Sie *srs.csaplan* aus.
- Klicken Sie auf Weiter.

Abbildung 22-42
Dialogfeld "Cox-Regression", Registerkarte "Zeit und Ereignis"



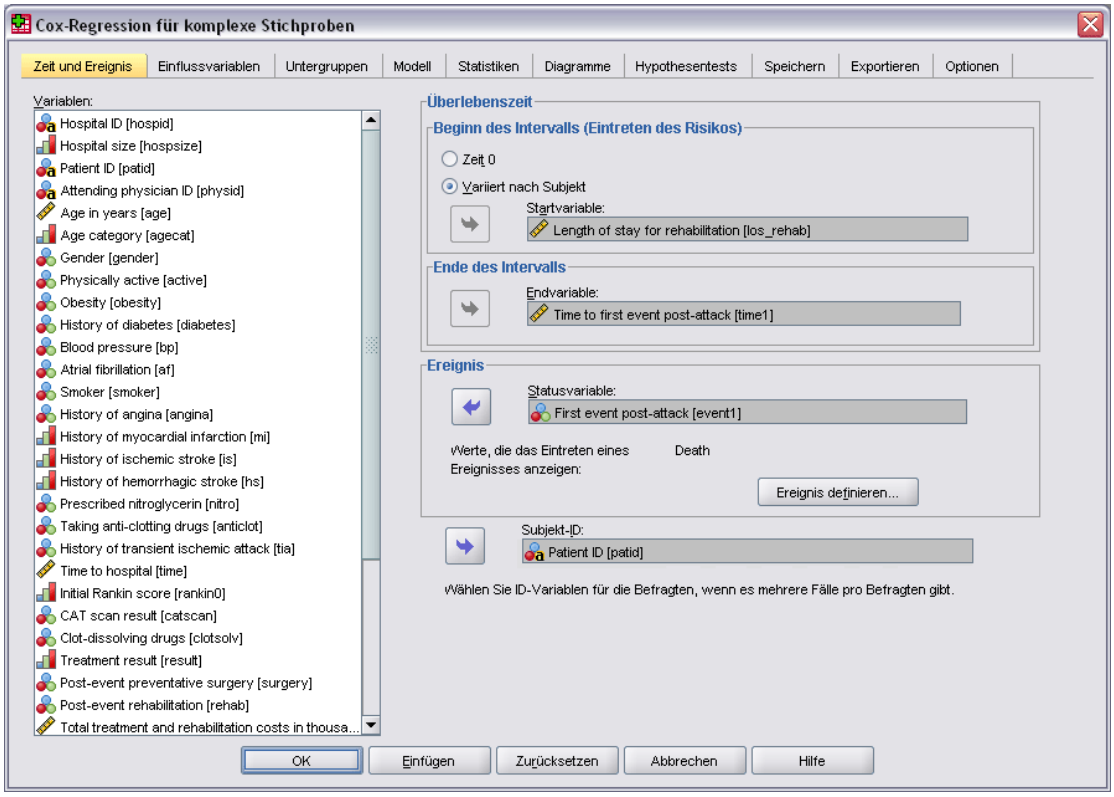
- Aktivieren Sie Variiert nach Subjekt und wählen Sie *Length of stay for rehabilitation [los_rehab]* (Aufenthaltsdauer für Rehabilitation) als Startvariable aus. Beachten Sie, dass die umstrukturierte Variable das Variablenlabel aus der ersten Variablen übernommen hat, die für ihre Erstellung verwendet wurde, obwohl das Label nicht unbedingt für die erstellte Variable geeignet ist.
- Wählen Sie *Time to first event post-attack [time_to_event]* (Zeit bis zum ersten Ereignis nach dem Anfall) als Endvariable aus.
- Wählen Sie *First event post-attack [event]* (Erstes Ereignis nach dem Anfall) als Statusvariable aus.
- Klicken Sie auf Ereignis definieren.

Abbildung 22-43
Dialogfeld "Ereignis definieren"



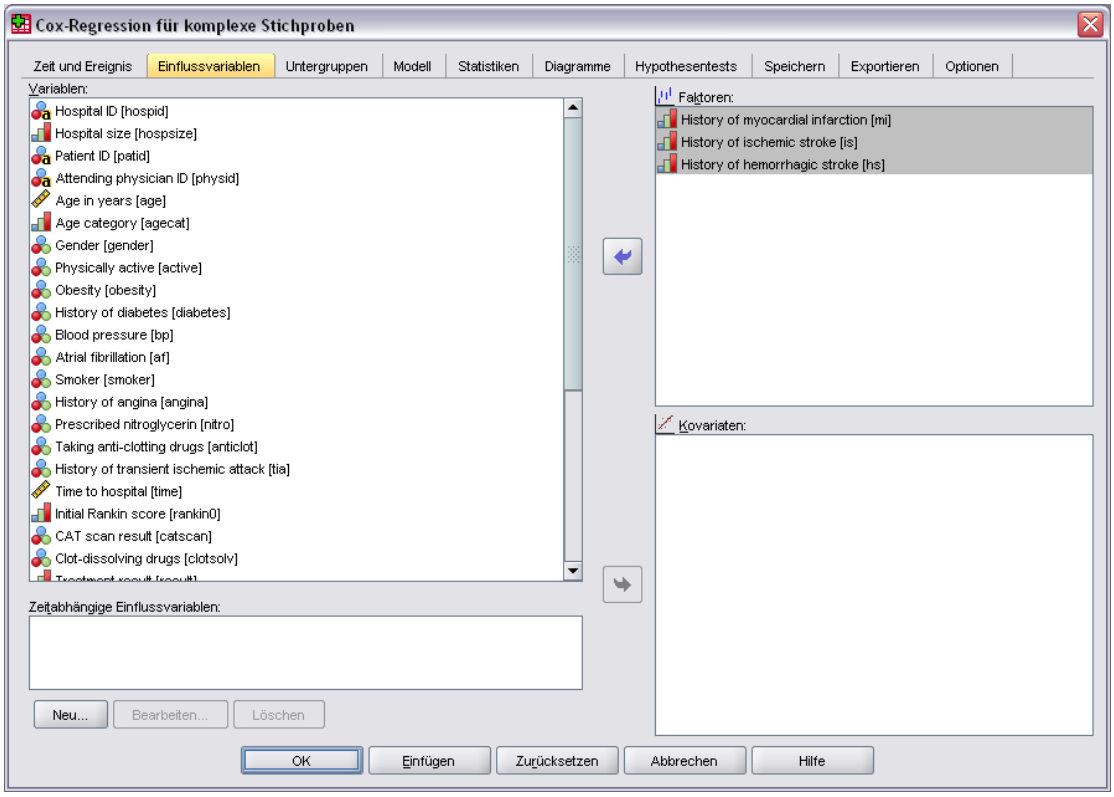
- Wählen Sie 4 Death (Tod) als Wert, der anzeigt, dass das terminale Ereignis eingetreten ist.
- Klicken Sie auf Weiter.

Abbildung 22-44
Dialogfeld "Cox-Regression", Registerkarte "Zeit und Ereignis"



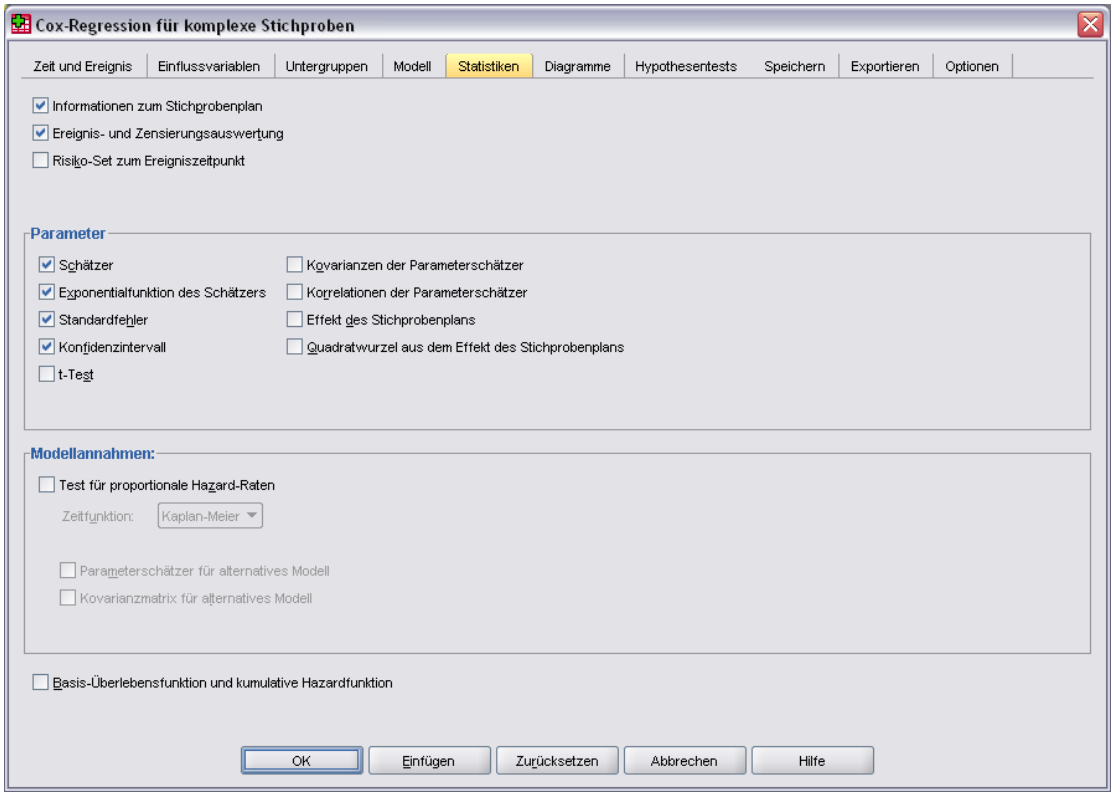
- Wählen Sie *Patient ID [patid]* (Patienten-ID) als Subjekt-ID aus.
- Klicken Sie auf die Registerkarte Einflussvariablen.

Abbildung 22-45
Dialogfeld "Cox-Regression", Registerkarte "Einflussvariablen"



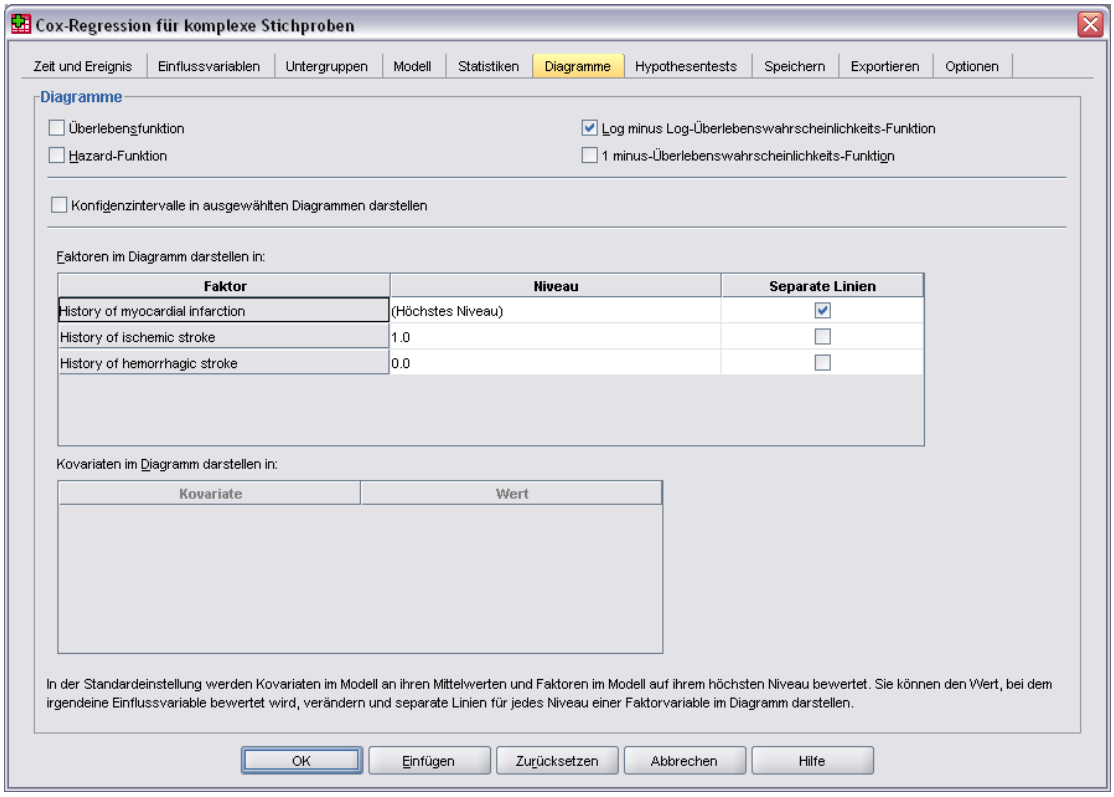
- Wählen Sie *History of myocardial infarction [mi]* (Herzinfarktanamnese) bis *History of hemorrhagic stroke [hs]* (hämorrhagische Schlaganfälle in der Anamnese) als Faktoren aus.
- Klicken Sie auf die Registerkarte Statistik.

Abbildung 22-46
Dialogfeld "Cox-Regression", Registerkarte "Statistiken"



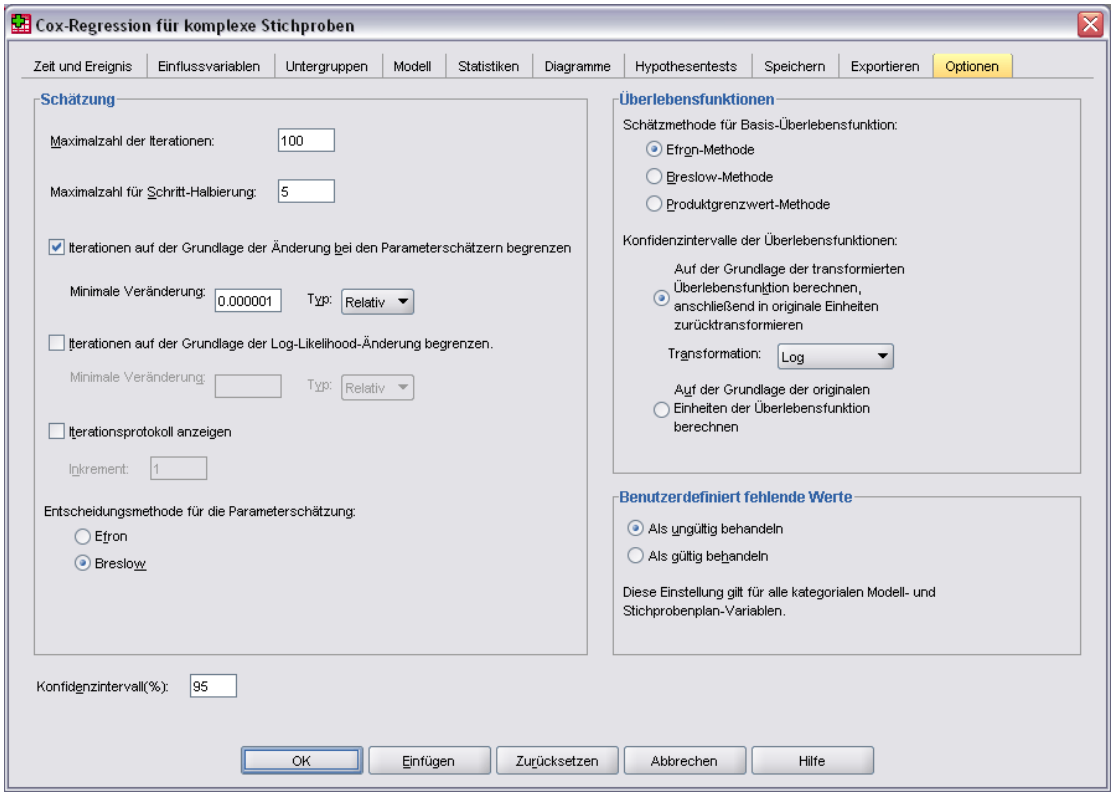
- Wählen Sie in der Gruppe "Parameter" die Optionen Schätzer, Exponentialfunktion des Schätzers, Standardfehler und Konfidenzintervall aus.
- Klicken Sie auf die Registerkarte Diagramme.

Abbildung 22-47
Dialogfeld "Cox-Regression", Registerkarte "Diagramme"



- Wählen Sie Log-minus-Log-Überlebenswahrscheinlichkeits-Funktion.
- Aktivieren Sie Separate Linien für *History of myocardial infarction* (Herzinfarktanamnese).
- Wählen Sie 1,0 als Stufe für *History of ischemic stroke* (ischämische Schlaganfälle in der Anamnese).
- Wählen Sie 0,0 als Stufe für *History of hemorrhagic stroke* (hämorrhagische Schlaganfälle in der Anamnese).
- Klicken Sie auf die Registerkarte Optionen.

Abbildung 22-48
Dialogfeld "Cox-Regression", Registerkarte "Optionen"



- Wählen Sie im Gruppenfeld "Schätzung" Breslow als Entscheidungsmethode.
- Klicken Sie auf OK.

Informationen zum Stichprobenplan

Abbildung 22-49
Informationen zum Stichprobenplan

		N
Gültig	Subjekte	2421
	Fälle	3310
	Ungültige Fälle	3953
	Gesamtzahl der Fälle	7263
	Subjektgröße bei Grundgesamtheit	2421,000
	Schichten	1
	Einheiten	2421
	Freiheitsgrade des Stichprobenplans	2420

Diese Tabelle enthält Informationen zu dem Stichprobenplan der zur Schätzung des Modells gehört.

- Es gibt für einige Subjekte mehrere Fälle und alle 3.310 Fälle werden in der Analyse verwendet.
- Der Plan enthält eine einzige Schicht und 2.421 Einheiten (eine für jedes Subjekt). Die Freiheitsgrade des Stichprobenplans werden als 2421−1=2420 geschätzt.

Tests der Modelleffekte

Abbildung 22-50
Tests der Modelleffekte

Quelle	Freiheitsgrade 1	Freiheitsgrade 2	Wald-F	Sig.
mi	3,000	2418,000	452,873	,000
is	2,000	2419,000	1064,936	,000
hs	2,000	2419,000	739,197	,000

Überlebenszeitvariable: Length of stay for rehabilitation, Time to first event post-attack
Ereignisstatusvariable: First event post-attack = 4
Subjekt-ID-Variable: Patient ID
Modell: mi, is, hs

Der Signifikanzwert für die einzelnen Effekte liegt nahe bei 0, was darauf hindeutet, dass sie alle zum Modell beitragen.

Parameter-Schätzer

Abbildung 22-51
Parameterschätzer

Parameter	B	Standardfehler	95%-Konfidenzintervall		Exp(B)	95% Konfidenzintervall für Exp (B)	
			Untere Grenze	Obere Grenze		Untere Grenze	Obere Grenze
[mi=0]	-6,381	,283	-6,935	-5,827	,002	,001	,003
[mi=1]	-5,589	,284	-6,147	-5,032	,004	,002	,007
[mi=2]	-2,119	,344	-2,794	-1,445	,120	,061	,236
[mi=3]	,000[a]	.	.	.	1,000	.	.
[is=1]	-6,421	,202	-6,817	-6,024	,002	,001	,002
[is=2]	-2,803	,222	-3,239	-2,366	,061	,039	,094
[is=3]	,000[a]	.	.	.	1,000	.	.
[hs=0]	-6,148	,355	-6,844	-5,453	,002	,001	,004
[hs=1]	-2,232	,373	-2,963	-1,502	,107	,052	,223
[hs=2]	,000[a]	.	.	.	1,000	.	.

Überlebenszeitvariable: Length of stay for rehabilitation, Time to first event post-attack
Ereignisstatusvariable: First event post-attack = 4
Subjekt-ID-Variable: Patient ID
Modell: mi, is, hs

a. Auf 0 gesetzt, da dieser Parameter redundant ist.

b. Entscheidungsmethode: Breslow

Die Prozedur verwendet die letzte Kategorie jedes Faktors als Referenzkategorie; der Effekt anderer Kategorien ist relativ zur Referenzkategorie. Hinweis: Der Schätzer ist zwar nützlich für statistische Tests, der potenzierte Schätzer, Exp(B), lässt sich jedoch leichter interpretieren, da die vorhergesagte Änderung in der Hazard-Rate relativ zur Referenzkategorie ist.

- Der Wert von Exp(B) für *[mi=0]* bedeutet, dass die Hazard-Rate für den Tod eines Patienten ohne frühere Herzinfarkte (mi) das 0,002-Fache der Hazard-Rate eines Patienten mit drei früheren Herzinfarkten beträgt.
- Die Konfidenzintervalle für *[mi=1]* und *[mi=0]* überschneiden sich, was darauf hindeutet, dass die Hazard-Rate für einen Patienten mit einem einzelnen früheren Herzinfarkt sich nicht in statistisch relevanter Weise von der eines Patienten ohne frühere Herzinfarkte unterscheidet.
- Die Konfidenzintervalle für *[mi=0]* und *[mi=1]* überschneiden sich nicht mit dem Intervall für *[mi=2]* und keines davon enthält den Wert 0. Dies deutet darauf hin, dass die Hazard-Rate für Patienten mit einem einzigen oder gar keinem früheren Herzinfarkt sich von der Hazard-Rate für Patienten mit zwei früheren Herzinfarkten unterscheidet, welche sich wiederum von der Hazard-Rate für Patienten mit drei früheren Herzinfarkten unterscheidet.

Ähnliche Beziehungen gelten für die Stufen von *is* und *hs*, bei denen die Erhöhung der Anzahl der früheren Vorfälle die Hazard-Rate für den Tod erhöht.

Musterwerte

Abbildung 22-52
Musterwerte

		Überlebenszeitintervall				
		Start	Ende	History of myocardial infarction	History of ischemic stroke	History of hemorrhagic stroke
Bezugsmuster	1	,000		Three	Three	Two
Muster 1.1	1	,000		None	One	None
Muster 1.2	1	,000		One	One	None
Muster 1.3	1	,000		Two	One	None
Muster 1.4	1	,000		Three	One	None

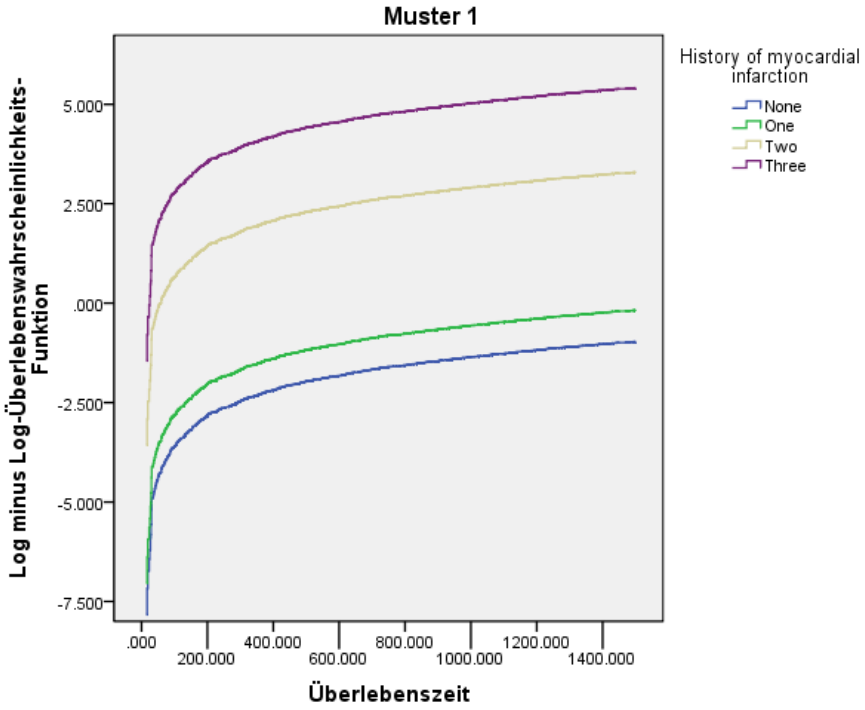
Einer nicht angegebenen Einflussvariabeln wird der Wert dieser Einflussvariablen beim Bezugsmuster zugewiesen.
Jedes Intervall für die Überlebenszeit ist als Start < Überlebenszeit <= Ende definiert.
Modell: mi, is, hs.

In der Tabelle mit den Musterwerten sind die Werte aufgelistet, die die Muster der einzelnen Einflussvariablen festlegen. Neben den Einflussvariablen im Modell werden die Start- und Endzeiten für das Überlebensintervall angezeigt. Bei Analysen, die über die Dialogfelder ausgeführt werden, ist die Startzeit immer 0 und die Endzeit immer unbegrenzt. Mithilfe von Syntax können Sie Pfade für stückweise konstante Einflussvariablen angeben.

- Das Bezugsmuster wird bei der Referenzkategorie der einzelnen Faktoren und beim Mittelwert der einzelnen Kovariaten festgelegt (in diesem Modell kommen keine Kovariaten vor). Bei diesem Daten-Set kann die Kombination der für das Referenzmodell angezeigten Faktoren nicht vorkommen; daher ignorieren wir das Log-minus-Log-Diagramm für das Bezugsmuster.
- Die Muster 1.1 bis 1.4 unterscheiden sich nur hinsichtlich des Werts von *History of myocardial infarction* (Herzinfarktanamnese). Für jeden Wert von *History of myocardial infarction* (Herzinfarktanamnese) wird ein separates Muster (und eine separate Linie im angeforderten Diagramm) erstellt, während die anderen Variablen konstant gehalten werden.

Log-Minus-Log-Diagramm

Abbildung 22-53
Log-Minus-Log-Diagramm



Dieses Diagramm zeigt die Log-minus-Log-Werte der Überlebensfunktion, ln(−ln(suvival)) in Abhängigkeit von der Überlebenszeit an. In diesem konkreten Diagramm wird eine separate Kurve für jede Kategorie von *History of myocardial infarction* (Herzinfarktanamnese) angezeigt, wobei *History of ischemic stroke* (ischämische Schlaganfälle in der Anamnese) beim Wert *One* (Einer) und *History of hemorrhagic stroke* (hämorrhagische Schlaganfälle in der Anamnese) beim Wert *None* (Keine) fixiert ist. Es bietet eine nützliche Visualisierung des Effekts von *History of myocardial infarction* (Herzinfarktanamnese) auf die Überlebensfunktion. Wie in der Tabelle der Parameterschätzer zu sehen, scheint die Überlebenszeit für Patienten mit einem einzigen oder gar keinem früheren Herzinfarkt sich von der Überlebenszeit für Patienten mit zwei früheren Herzinfarkten zu unterscheiden, welche sich wiederum von der Überlebenszeit für Patienten mit drei früheren Herzinfarkten unterscheidet.

Zusammenfassung

Sie haben ein Cox-Regressionsmodell für die Überlebenszeit nach Schlaganfällen angepasst, das die Effekte der Veränderung der Patientengeschichte nach dem Schlaganfall schätzt. Dies ist natürlich nur ein Anfang, da die Forscher zweifellos bestrebt wären, noch andere potenzielle Einflussvariablen in das Modell aufzunehmen. Außerdem könnten Sie in einer weiteren Analyse dieses Daten-Sets signifikantere Änderungen an der Modellstruktur betrachten. Beispielsweise geht das aktuelle Modell davon aus, dass der Effekt eines Ereignisses, das die Patientenanamnese

verändert, sich durch Anwendung eines Multiplikators auf die Basis-Hazard-Rate quantifizieren lässt. Stattdessen kann es sinnvoll sein anzunehmen, dass sich durch das eintreten eines nichttödlichen Ereignisses die Form der Basis-Hazard-Rate ändert. Um dies zu erreichen, könnten Sie die Analyse auf der Grundlage von *Event index* (Ereignisvariable) schichten.

Anhang

A

Beispieldateien

Die zusammen mit dem Produkt installierten Beispieldateien finden Sie im Unterverzeichnis *Samples* des Installationsverzeichnisses.

Beschreibungen

Im Folgenden finden Sie Kurzbeschreibungen der in den verschiedenen Beispielen in der Dokumentation verwendeten Beispieldateien:

- **accidents.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die alters- und geschlechtsabhängige Risikofaktoren für Autounfälle in einer bestimmten Region untersucht. Jeder Fall entspricht einer Kreuzklassifikation von Alterskategorie und Geschlecht.
- **adl.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bemühungen geht, die Vorteile einer vorgeschlagenen Therapieform für Schlaganfallpatienten zu ermitteln. Ärzte teilten weibliche Schlaganfallpatienten nach dem Zufallsprinzip jeweils einer von zwei Gruppen zu. Die erste Gruppe erhielt die physische Standardtherapie, die zweite erhielt eine zusätzliche Emotionaltherapie. Drei Monate nach den Behandlungen wurden die Fähigkeiten der einzelnen Patienten, übliche Alltagsaktivitäten auszuführen, als ordinale Variablen bewertet.
- **advert.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Einzelhändlers geht, die Beziehungen zwischen den in Werbung investierten Beträgen und den daraus resultierenden Umsätzen zu untersuchen. Zu diesem Zweck hat er die Umsätze vergangener Jahre und die zugehörigen Werbeausgaben zusammengestellt.
- **aflatoxin.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Tests von Maisernten auf Aflatoxin geht, ein Gift, dessen Konzentration stark zwischen und innerhalb von Ernteerträgen schwankt. Ein Kornverarbeitungsbetrieb hat aus 8 Ernteerträgen je 16 Proben erhalten und das Aflatoxinniveau in Teilen pro Milliarde (parts per billion, PPB) gemessen.
- **aflatoxin20.sav.** Diese Datendatei enthält die Aflatoxinmessungen aus jeder der 16 Stichproben aus den Erträgen 4 und 8 der Datendatei *aflatoxin.sav*.
- **anorectic.sav.** Bei der Ausarbeitung einer standardisierten Symptomatologie anorektischen/bulimischen Verhaltens führten Forscher (Van der Ham, Meulman, Van Strien, als auch Van Engeland, 1997) eine Studie mit 55 Jugendlichen mit bekannten Ess-Störungen durch. Jeder Patient wurde vier Mal über einen Zeitraum von vier Jahren untersucht, es fanden also insgesamt 220 Beobachtungen statt. Bei jeder Beobachtung erhielten die Patienten Scores für jedes von 16 Symptomen. Die Symptomwerte fehlen für Patient 71

zum Zeitpunkt 2, Patient 76 zum Zeitpunkt 2 und Patient 47 zum Zeitpunkt 3, wodurch 217 gültige Beobachtungen verbleiben.

- **autoaccidents.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Versicherungsanalysten geht, ein Modell zur Anzahl der Autounfälle pro Fahrer unter Berücksichtigung von Alter und Geschlecht zu erstellen. Jeder Fall stellt einen Fahrer dar und erfasst das Geschlecht des Fahrers, sein Alter in Jahren und die Anzahl der Autounfälle in den letzten fünf Jahren.
- **band.sav.** Diese Datendatei enthält die hypothetischen wöchentlichen Verkaufszahlen von CDs für eine Musikgruppe. Daten für drei mögliche Einflussvariablen wurden ebenfalls aufgenommen.
- **bankloan.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Bank geht, den Anteil der nicht zurückgezahlten Kredite zu reduzieren. Die Datei enthält Informationen zum Finanzstatus und demografischen Hintergrund von 850 früheren und potenziellen Kunden. Bei den ersten 700 Fällen handelt es sich um Kunden, denen bereits ein Kredit gewährt wurde. Bei den letzten 150 Fällen handelt es sich um potenzielle Kunden, deren Kreditrisiko die Bank als gering oder hoch einstufen möchte.
- **bankloan_binning.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Informationen zum Finanzstatus und demografischen Hintergrund von 5.000 früheren Kunden enthält.
- **behavior.sav.** In einem klassischen Beispiel (Price als auch Bouffard, 1974) wurden 52 Schüler/Studenten gebeten, die Kombinationen aus 15 Situationen und 15 Verhaltensweisen auf einer 10-Punkte-Skala von 0 = "ausgesprochen angemessen" bis 9 = "ausgesprochen unangemessen" zu bewerten. Die Werte werden über die einzelnen Personen gemittelt und als Unähnlichkeiten verwendet.
- **behavior_ini.sav.** Diese Datendatei enthält eine Ausgangskonfiguration für eine zweidimensionale Lösung für *behavior.sav*.
- **brakes.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Fabrik geht, die Scheibenbremsen für Hochleistungsautomobile herstellt. Die Datendatei enthält Messungen des Durchmessers von 16 Scheiben aus 8 Produktionsmaschinen. Der Zieldurchmesser für die Scheiben ist 322 Millimeter.
- **breakfast.sav.** In einer klassischen Studie (Green als auch Rao, 1972) wurden 21 MBA-Studenten der Wharton School mit ihren Lebensgefährten darum gebeten, 15 Frühstücksartikel in der Vorzugsreihenfolge von 1 = "am meisten bevorzugt" bis 15 = "am wenigsten bevorzugt" zu ordnen. Die Bevorzugungen wurden in sechs unterschiedlichen Szenarien erfasst, von "Overall preference" (Allgemein bevorzugt) bis "Snack, with beverage only" (Imbiss, nur mit Getränk).
- **breakfast-overall.sav.** Diese Datei enthält die Daten zu den bevorzugten Frühstücksartikeln, allerdings nur für das erste Szenario, "Overall preference" (Allgemein bevorzugt).
- **broadband_1.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die die Anzahl der Abonnenten eines Breitband-Service, nach Region geordnet, enthält. Die Datendatei enthält die monatlichen Abonnentenzahlen für 85 Regionen über einen Zeitraum von vier Jahren.
- **broadband_2.sav.** Diese Datendatei stimmt mit *broadband_1.sav* überein, enthält jedoch Daten für weitere drei Monate.

- **car_insurance_claims.sav.** Ein an anderer Stelle (McCullagh als auch Nelder, 1989) vorgestelltes und analysiertes Daten-Set bezieht sich auf Schadensansprüche für Autos. Die durchschnittliche Höhe der Schadensansprüche lässt sich mit Gamma-Verteilung modellieren. Dazu wird eine inverse Link-Funktion verwendet, um den Mittelwert der abhängigen Variablen mit einer linearen Kombination aus Alter des Versicherungsnehmers, Fahrzeugtyp und Fahrzeugalter in Bezug zu setzen. Die Anzahl der eingereichten Schadensansprüche kann als Skalierungsgewicht verwendet werden.
- **car_sales.sav.** Diese Datendatei enthält hypothetische Verkaufsschätzer, Listenpreise und physische Spezifikationen für verschiedene Fahrzeugfabrikate und -modelle. Die Listenpreise und physischen Spezifikationen wurden von *edmunds.com* und Hersteller-Websites entnommen.
- **carpet.sav.** In einem beliebten Beispiel möchte (Green als auch Wind, 1973) einen neuen Teppichreiniger vermarkten und dazu den Einfluss von fünf Faktoren auf die Bevorzugung durch den Verbraucher untersuchen: Verpackungsgestaltung, Markenname, Preis, Gütesiegel, *Good Housekeeping* und Geld-zurück-Garantie. Die Verpackungsgestaltung liegt in drei Faktorstufen vor, die sich durch die Position der Auftragebürste unterscheiden. Außerdem gibt es drei Markennamen (*K2R*, *Glory* und *Bissell*), drei Preisstufen sowie je zwei Stufen (Nein oder Ja) für die letzten beiden Faktoren. 10 Kunden stufen 22 Profile ein, die durch diese Faktoren definiert sind. Die Variable *Preference* enthält den Rang der durchschnittlichen Einstufung für die verschiedenen Profile. Ein niedriger Rang bedeutet eine starke Bevorzugung. Diese Variable gibt ein Gesamtmaß der Bevorzugung für die Profile an.
- **carpet_prefs.sav.** Diese Datendatei beruht auf denselben Beispielen, wie für *carpet.sav* beschrieben, enthält jedoch die tatsächlichen Einstufungen durch jeden der 10 Kunden. Die Kunden wurden gebeten, die 22 Produktprofile in der Reihenfolge ihrer Präferenzen einzustufen. Die Variablen *PREF1* bis *PREF22* enthalten die IDs der zugeordneten Profile, wie in *carpet_plan.sav* definiert.
- **catalog.sav.** Diese Datendatei enthält hypothetische monatliche Verkaufszahlen für drei Produkte, die von einem Versandhaus verkauft werden. Daten für fünf mögliche Einflussvariablen wurden ebenfalls aufgenommen.
- **catalog_seasfac.sav.** Diese Datendatei ist mit *catalog.sav* identisch, außer, dass ein Set von saisonalen Faktoren, die mithilfe der Prozedur "Saisonale Zerlegung" berechnet wurden, sowie die zugehörigen Datumsvariablen hinzugefügt wurden.
- **cellular.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Mobiltelefonunternehmens geht, die Kundenabwanderung zu verringern. Scores für die Abwanderungsneigung (von 0 bis 100) werden auf die Kunden angewendet. Kunden mit einem Score von 50 oder höher streben vermutlich einen Anbieterwechsel an.
- **ceramics.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Herstellers geht, der ermitteln möchte, ob ein neue, hochwertige Keramiklegierung eine größere Hitzebeständigkeit aufweist als eine Standardlegierung. Jeder Fall entspricht einem Test einer der Legierungen; die Temperatur, bei der das Keramikwälzlager versagte, wurde erfasst.
- **cereal.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Umfrage geht, bei der 880 Personen nach ihren Frühstückgewohnheiten befragt wurden. Außerdem wurden Alter, Geschlecht, Familienstand und Vorliegen bzw. Nichtvorliegen eines aktiven Lebensstils (auf der Grundlage von mindestens zwei Trainingseinheiten pro Woche) erfasst. Jeder Fall entspricht einem Teilnehmer.

- **clothing_defects.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Bekleidungsfabrik geht. Aus jeder in der Fabrik produzierten Charge entnehmen die Kontrolleure eine Stichprobe an Bekleidungsartikeln und zählen die Anzahl der Bekleidungsartikel die inakzeptabel sind.
- **coffee.sav.** Diese Datendatei enthält Daten zum wahrgenommenen Image von sechs Eiskaffeemarken (Kennedy, Riquier, als auch Sharp, 1996). Bei den 23 Attributen des Eiskaffee-Image sollten die Teilnehmer jeweils alle Marken auswählen, die durch dieses Attribut beschrieben werden. Die sechs Marken werden als "AA", "BB", "CC", "DD", "EE" und "FF" bezeichnet, um Vertraulichkeit zu gewährleisten.
- **contacts.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Kontaktlisten einer Gruppe von Vertretern geht, die Computer an Unternehmen verkaufen. Die einzelnen Kontaktpersonen werden anhand der Abteilung, in der sie in ihrem Unternehmen arbeiten und anhand ihrer Stellung in der Unternehmenshierarchie in Kategorien eingeteilt. Außerdem werden der Betrag des letzten Verkaufs, die Zeit seit dem letzten Verkauf und die Größe des Unternehmens, in dem die Kontaktperson arbeitet, aufgezeichnet.
- **creditpromo.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Kaufhauses geht, die Wirksamkeit einer kürzlich durchgeführten Kreditkarten-Werbeaktion einzuschätzen. Dazu wurden 500 Karteninhaber nach dem Zufallsprinzip ausgewählt. Die Hälfte erhielt eine Werbebeilage, die einen reduzierten Zinssatz für Einkäufe in den nächsten drei Monaten ankündigte. Die andere Hälfte erhielt eine Standard-Werbebeilage.
- **customer_dbase.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Unternehmens geht, das die Informationen in seinem Data Warehouse nutzen möchte, um spezielle Angebote für Kunden zu erstellen, die mit der größten Wahrscheinlichkeit darauf ansprechen. Nach dem Zufallsprinzip wurde eine Untergruppe des Kundenstamms ausgewählt. Diese Gruppe erhielt die speziellen Angebote und die Reaktionen wurden aufgezeichnet.
- **customers_model.sav.** Diese Datei enthält hypothetische Daten zu Einzelpersonen, auf die sich eine Marketingkampagne richtete. Zu diesen Daten gehören demografische Informationen, eine Übersicht über die bisherigen Einkäufe und die Angabe ob die einzelnen Personen auf die Kampagne ansprachen oder nicht. Jeder Fall entspricht einer Einzelperson.
- **customers_new.sav.** Diese Datei enthält hypothetische Daten zu Einzelpersonen, die potenzielle Kandidaten für Marketingkampagnen sind. Zu diesen Daten gehören demografische Informationen und eine Übersicht über die bisherigen Einkäufe für jede Person. Jeder Fall entspricht einer Einzelperson.
- **debate.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die gepaarte Antworten auf eine Umfrage unter den Zuhörern einer politischen Debatte enthält (Antworten vor und nach der Debatte). Jeder Fall entspricht einem Befragten.
- **debate_aggregate.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der die Antworten aus *debate.sav* aggregiert wurden. Jeder Fall entspricht einer Kreuzklassifikation der bevorzugten Politiker vor und nach der Debatte.
- **demo.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Kundendatenbank geht, die zum Zwecke der Zusendung monatlicher Angebote erworben wurde. Neben verschiedenen demografischen Informationen ist erfasst, ob der Kunde auf das Angebot geantwortet hat.

- **demo_cs_1.sav.** Hierbei handelt es sich um eine hypothetische Datendatei für den ersten Schritt eines Unternehmens, das eine Datenbank mit Umfrageinformationen zusammenstellen möchte. Jeder Fall entspricht einer anderen Stadt. Außerdem sind IDs für Region, Provinz, Landkreis und Stadt erfasst.
- **demo_cs_2.sav.** Hierbei handelt es sich um eine hypothetische Datendatei für den zweiten Schritt eines Unternehmens, das eine Datenbank mit Umfrageinformationen zusammenstellen möchte. Jeder Fall entspricht einem anderen Stadtteil aus den im ersten Schritt ausgewählten Städten. Außerdem sind IDs für Region, Provinz, Landkreis, Stadt, Stadtteil und Wohneinheit erfasst. Die Informationen zur Stichprobenziehung aus den ersten beiden Stufen des Stichprobenplans sind ebenfalls enthalten.
- **demo_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Umfrageinformationen enthält die mit einem komplexen Stichprobenplan erfasst wurden. Jeder Fall entspricht einer anderen Wohneinheit. Es sind verschiedene Informationen zum demografischen Hintergrund und zur Stichprobenziehung erfasst.
- **dietstudy.sav.** Diese hypothetische Datendatei enthält die Ergebnisse einer Studie der "Stillman-Diät" (Rickman, Mitchell, Dingman, als auch Dalen, 1974). Jeder Fall entspricht einem Teilnehmer und enthält dessen Gewicht vor und nach der Diät in amerikanischen Pfund sowie mehrere Messungen des Triglyceridspiegels (in mg/100 ml).
- **dischargedata.sav.** Hierbei handelt es sich um eine Datendatei zum Thema *Seasonal Patterns of Winnipeg Hospital Use*, (Menec , Roos, Nowicki, MacWilliam, Finlayson , als auch Black, 1999) (Saisonale Muster der Belegung im Krankenhaus von Winnipeg) vom Manitoba Centre for Health Policy.
- **dvdplayer.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Entwicklung eines neuen DVD-Spielers geht. Mithilfe eines Prototyps hat das Marketing-Team Zielgruppendaten erfasst. Jeder Fall entspricht einem befragten Benutzer und enthält demografische Daten zu dem Benutzer sowie dessen Antworten auf Fragen zum Prototyp.
- **flying.sav.** Diese Datendatei enthält die Flugmeilen zwischen zehn Städten in den USA.
- **german_credit.sav.** Diese Daten sind aus dem Daten-Set "German credit" im Repository of Machine Learning Databases (Blake als auch Merz, 1998) an der Universität von Kalifornien in Irvine entnommen.
- **grocery_1month.sav.** Bei dieser hypothetischen Datendatei handelt es sich um die Datendatei *grocery_coupons.sav*, wobei die wöchentlichen Einkäufe zusammengefasst sind, sodass jeder Fall einem anderen Kunden entspricht. Dadurch entfallen einige der Variablen, die wöchentlichen Änderungen unterworfen waren, und der verzeichnete ausgegebene Betrag ist nun die Summe der Beträge, die in den vier Wochen der Studie ausgegeben wurden.
- **grocery_coupons.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Umfragedaten enthält, die von einer Lebensmittelkette erfasst wurden, die sich für die Kaufgewohnheiten ihrer Kunden interessiert. Jeder Kunde wird über vier Wochen beobachtet, und jeder Fall entspricht einer Kundenwoche und enthält Informationen zu den Geschäften, in denen der Kunde einkauft sowie zu anderen Merkmalen, beispielsweise welcher Betrag in der betreffenden Woche für Lebensmittel ausgegeben wurde.
- **guttman.sav.** Bell (Bell, 1961) legte eine Tabelle zur Darstellung möglicher sozialer Gruppen vor. Guttman (Guttman, 1968) verwendete einen Teil dieser Tabelle, bei der fünf Variablen, die Aspekte beschreiben, wie soziale Interaktion, das Gefühl der Gruppenzugehörigkeit, die

physische Nähe der Mitglieder und die Formalität der Beziehung, mit sieben theoretischen sozialen Gruppen gekreuzt wurden: “crowds” (Menschenmassen, beispielsweise die Zuschauer eines Fußballspiels), “audience” (Zuhörerschaften, beispielsweise die Personen im Theater oder bei einer Vorlesung), “public” (Öffentlichkeit, beispielsweise Zeitungsleser oder Fernsehzuschauer), “mobs” (Mobs, wie Menschenmassen, jedoch mit wesentlich stärkerer Interaktion), “primary groups” (Primärgruppen, vertraulich), “secondary groups” (Sekundärgruppen, freiwillig) und “modern community” (die moderne Gesellschaft, ein lockerer Zusammenschluss, der aus einer engen physischen Nähe und dem Bedarf an spezialisierten Dienstleistungen entsteht).

- **healthplans.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Versicherungsgruppe geht, vier verschiedene Pläne zur Gesundheitsvorsorge für Kleinbetriebe zu evaluieren. Zwölf Inhaber von Kleinbetrieben (Arbeitgeber) wurden gebeten, die Pläne danach in eine Rangfolge zu bringen, wie gern sie sie ihren Mitarbeitern anbieten würden. Jeder Fall entspricht einem Arbeitgeber und enthält die Reaktionen auf die einzelnen Pläne.
- **health_funding.sav.** Hierbei handelt es sich um eine hypothetische Datei, die Daten zur Finanzierung des Gesundheitswesens (Betrag pro 100 Personen), Krankheitsraten (Rate pro 10.000 Personen der Bevölkerung) und Besuche bei medizinischen Einrichtungen/Ärzten (Rate pro 10.000 Personen der Bevölkerung) enthält. Jeder Fall entspricht einer anderen Stadt.
- **hivassay.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu den Bemühungen eines pharmazeutischen Labors, einen Schnelltest zur Erkennung von HIV-Infektionen zu entwickeln. Die Ergebnisse des Tests sind acht kräftiger werdende Rotschattierungen, wobei kräftigeren Schattierungen auf eine höhere Infektionswahrscheinlichkeit hindeuten. Bei 2.000 Blutproben, von denen die Hälfte mit HIV infiziert war, wurde ein Labortest durchgeführt.
- **hourlywagedata.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zum Stundenlohn von Pflegepersonal in Praxen und Krankenhäusern mit unterschiedlich langer Berufserfahrung.
- **insure.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die die Risikofaktoren untersucht, die darauf hinweisen, ob ein Kunde die Leistungen einer mit einer Laufzeit von 10 Jahren abgeschlossenen Lebensversicherung in Anspruch nehmen wird. Jeder Fall in der Datendatei entspricht einem Paar von Verträgen, je einer mit Leistungsforderung und der andere ohne, wobei die beiden Versicherungsnehmer in Alter und Geschlecht übereinstimmen.
- **judges.sav.** Hierbei handelt es sich um eine hypothetische Datendatei mit den Wertungen von ausgebildeten Kampfrichtern (sowie eines Sportliebhabers) zu 300 Kunstturnleistungen. Jede Zeile stellt eine Leistung dar; die Kampfrichter bewerteten jeweils dieselben Leistungen.
- **kinship_dat.sav.** Rosenberg und Kim (Rosenberg als auch Kim, 1975) haben 15 Bezeichnungen für den Verwandtschaftsgrad untersucht (Tante, Bruder, Cousin, Tochter, Vater, Enkelin, Großvater, Großmutter, Enkel, Mutter, Neffe, Nichte, Schwester, Sohn, Onkel). Die beiden Analytiker baten vier Gruppen von College-Studenten (zwei weibliche und zwei männliche Gruppen), diese Bezeichnungen auf der Grundlage der Ähnlichkeiten zu sortieren. Zwei Gruppen (eine weibliche und eine männliche Gruppe) wurden gebeten, die Bezeichnungen zweimal zu sortieren; die zweite Sortierung sollte dabei nach einem anderen Kriterium erfolgen als die erste. So wurden insgesamt sechs “Quellen” erzielt. Jede Quelle entspricht einer Ähnlichkeitsmatrix mit 15×15 Elementen. Die Anzahl der Zellen ist dabei gleich der

Anzahl der Personen in einer Quelle minus der Anzahl der gemeinsamen Platzierungen der Objekte in dieser Quelle.

- **kinship_ini.sav.** Diese Datendatei enthält eine Ausgangskonfiguration für eine dreidimensionale Lösung für *kinship_dat.sav*.
- **kinship_var.sav.** Diese Datendatei enthält die unabhängigen Variablen *gender* (Geschlecht), *gener* (Generation) und *degree* (Verwandtschaftsgrad), die zur Interpretation der Dimensionen einer Lösung für *kinship_dat.sav* verwendet werden können. Insbesondere können sie verwendet werden, um den Lösungsraum auf eine lineare Kombination dieser Variablen zu beschränken.
- **mailresponse.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines Bekleidungsherstellers geht, der ermitteln möchte, ob die Verwendung von Briefsendungen für das Direktmarketing zu schnelleren Antworten führt als Postwurfsendungen. Die Mitarbeiter in der Bestellannahme erfassen, wie vielen Wochen nach der Postsendung die einzelnen Bestellungen aufgegeben wurden.
- **marketvalues.sav.** Diese Datendatei betrifft Hausverkäufe in einem Neubaugebiet in Algonquin, Illinois, in den Jahren 1999–2000. Diese Verkäufe sind in Grundbucheinträgen dokumentiert.
- **mutualfund.sav.** Diese Datendatei betrifft Aktienmarktdaten für verschiedene Technologieaktien, die in im Index S&P 500 verzeichnet sind. Jeder Fall entspricht einem Unternehmen.
- **nhis2000_subset.sav.** Die "National Health Interview Survey (NHIS)" ist eine große, bevölkerungsbezogene Umfrage in unter der US-amerikanischen Zivilbevölkerung. Es werden persönliche Interviews in einer landesweit repräsentativen Stichprobe von Haushalten durchgeführt. Für die Mitglieder jedes Haushalts werden demografische Informationen und Beobachtungen zum Gesundheitsverhalten und Gesundheitsstatus eingeholt. Diese Datendatei enthält eine Teilmenge der Informationen aus der Umfrage des Jahres 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Datendatei und Dokumentation öffentlich zugänglich. *ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/*. Zugriff erfolgte 2003.
- **ozone.sav.** Die Daten enthalten 330 Beobachtungen zu sechs meteorologischen Variablen zur Vorhersage der Ozonkonzentration aus den übrigen Variablen. Bei früheren Untersuchungen (Breiman als auch Friedman, 1985), (Hastie als auch Tibshirani, 1990) fanden Wissenschaftler einige Nichtlinearitäten unter diesen Variablen, die die Standardverfahren bei der Regression behindern.
- **pain_medication.sav.** Diese hypothetische Datendatei enthält die Ergebnisse eines klinischen Tests für ein entzündungshemmendes Medikament zur Schmerzbehandlung bei chronischer Arthritis. Von besonderem Interesse ist die Zeitdauer, bis die Wirkung des Medikaments einsetzt und wie es im Vergleich mit bestehenden Medikamenten abschneidet.
- **patient_los.sav.** Diese hypothetische Datendatei enthält die Behandlungsaufzeichnungen zu Patienten, die wegen des Verdachts auf Herzinfarkt in das Krankenhaus eingeliefert wurden. Jeder Fall entspricht einem Patienten und enthält diverse Variablen in Bezug auf den Krankenhausaufenthalt.

- **patlos_sample.sav.** Diese hypothetische Datendatei enthält die Behandlungsaufzeichnungen für eine Stichprobe von Patienten, denen während der Behandlung eines Herzinfarkts Thrombolytika verabreicht wurden. Jeder Fall entspricht einem Patienten und enthält diverse Variablen in Bezug auf den Krankenhausaufenthalt.
- **polishing.sav.** Hierbei handelt es sich um die Datendatei "Nambeware Polishing Times" aus der Data and Story Library. Sie bezieht sich auf die Bemühungen eines Herstellers von Metallgeschirr (Nambe Mills, Santa Fe, New Mexico) zur zeitlichen Planung seiner Produktion. Jeder Fall entspricht einem anderen Artikel in der Produktpalette. Für jeden Artikel sind Durchmesser, Polierzeit, Preis und Produkttyp erfasst.
- **poll_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bemühungen geht, die öffentliche Unterstützung für einen Gesetzentwurf zu ermitteln, bevor er im Parlament eingebracht wird. Die Fälle entsprechen registrierten Wählern. Für jeden Fall sind County, Gemeinde und Wohnviertel des Wählers erfasst.
- **poll_cs_sample.sav.** Diese hypothetische Datendatei enthält eine Stichprobe der in *poll_cs.sav* aufgeführten Wähler. Die Stichprobe wurde gemäß dem in der Plandatei *poll.csplan* angegebenen Stichprobenplan gezogen und in dieser Datendatei sind die Einschlusswahrscheinlichkeiten und Stichprobengewichtungen erfasst. Beachten Sie jedoch Folgendes: Da im Stichprobenplan die PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*poll_jointprob.sav*). Die zusätzlichen Variablen zum demografischen Hintergrund der Wähler und ihrer Meinung zum vorgeschlagenen Gesetzentwurf wurden nach der Ziehung der Stichprobe erfasst und zur Datendatei hinzugefügt.
- **property_assess.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines für einen Bezirk (County) zuständigen Immobilienbewerters geht, trotz eingeschränkter Ressourcen die Einschätzungen des Werts von Immobilien auf dem aktuellsten Stand zu halten. Die Fälle entsprechen den Immobilien, die im vergangenen Jahr in dem betreffenden County verkauft wurden. Jeder Fall in der Datendatei enthält die Gemeinde, in der sich die Immobilie befindet, den Bewerter, der die Immobilie besichtigt hat, die seit dieser Bewertung verstrichene Zeit, den zu diesem Zeitpunkt ermittelten Wert sowie den Verkaufswert der Immobilie.
- **property_assess_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines für einen US-Bundesstaat zuständigen Immobilienbewerters geht, trotz eingeschränkter Ressourcen die Einschätzungen des Werts von Immobilien auf dem aktuellsten Stand zu halten. Die Fälle entsprechen den Immobilien in dem betreffenden Bundesstaat. Jeder Fall in der Datendatei enthält das County, die Gemeinde und das Wohnviertel, in dem sich die Immobilie befindet, die seit der letzten Bewertung verstrichene Zeit sowie zu diesem Zeitpunkt ermittelten Wert.
- **property_assess_cs_sample.sav.** Diese hypothetische Datendatei enthält eine Stichprobe der in *property_assess_cs.sav* aufgeführten Immobilien. Die Stichprobe wurde gemäß dem in der Plandatei *property_assess.csplan* angegebenen Stichprobenplan gezogen und in dieser Datendatei sind die Einschlusswahrscheinlichkeiten und Stichprobengewichtungen erfasst. Die zusätzliche Variable *Current value* (Aktueller Wert) wurde nach der Ziehung der Stichprobe erfasst und zur Datendatei hinzugefügt.

- **recidivism.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Strafverfolgungsbehörde geht, einen Einblick in die Rückfallraten in ihrem Zuständigkeitsbereich zu gewinnen. Jeder Fall entspricht einem frühren Straftäter und erfasst Daten zu dessen demografischen Hintergrund, einige Details zu seinem ersten Verbrechen sowie die Zeit bis zu seiner zweiten Festnahme, sofern diese innerhalb von zwei Jahren nach der ersten Festnahme erfolgte.
- **recidivism_cs_sample.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Strafverfolgungsbehörde geht, einen Einblick in die Rückfallraten in ihrem Zuständigkeitsbereich zu gewinnen. Jeder Fall entspricht einem früheren Straftäter, der im Juni 2003 erstmals aus der Haft entlassen wurde, und erfasst Daten zu dessen demografischen Hintergrund, einige Details zu seinem ersten Verbrechen sowie die Daten zu seiner zweiten Festnahme, sofern diese bis Ende Juni 2006 erfolgte. Die Straftäter wurden aus per Stichprobenziehung ermittelten Polizeidirektionen ausgewählt (gemäß dem in *recidivism_cs.csplan* angegebenen Stichprobenplan). Da hierbei eine PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*recidivism_cs_jointprob.sav*).
- **salesperformance.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bewertung von zwei neuen Verkaufsschulungen geht. 60 Mitarbeiter, die in drei Gruppen unterteilt sind, erhalten jeweils eine Standardschulung. Zusätzlich erhält Gruppe 2 eine technische Schulung und Gruppe 3 eine Praxisschulung. Die einzelnen Mitarbeiter wurden am Ende der Schulung einem Test unterzogen und die erzielten Punkte wurden erfasst. Jeder Fall in der Datendatei stellt einen Lehrgangsteilnehmer dar und enthält die Gruppe, der der Lehrgangsteilnehmer zugeteilt wurde sowie die von ihm in der Prüfung erreichte Punktzahl.
- **satisf.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einer Zufriedenheitsumfrage, die von einem Einzelhandelsunternehmen in 4 Filialen durchgeführt wurde. Insgesamt wurden 582 Kunden befragt. Jeder Fall gibt die Antworten eines einzelnen Kunden wieder.
- **screws.sav.** Diese Datendatei enthält Informationen zu den Eigenschaften von Schrauben, Bolzen, Muttern und Reißnägeln (Hartigan, 1975).
- **shampoo_ph.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Fabrik für Haarpflegeprodukte geht. In regelmäßigen Zeitabständen werden Messwerte von sechs separaten Ausgangschargen erhoben und ihr pH-Wert erfasst. Der Zielbereich ist 4,5–5,5.
- **ships.sav.** Ein an anderer Stelle (McCullagh et al., 1989) vorgestelltes und analysiertes Daten-Set bezieht sich auf die durch Wellen verursachten Schäden an Frachtschiffen. Die Vorfallshäufigkeiten können unter Angabe von Schiffstyp, Konstruktionszeitraum und Betriebszeitraum gemäß einer Poisson-Rate modelliert werden. Das Aggregat der Betriebsmonate für jede Zelle der durch die Kreuzklassifizierung der Faktoren gebildeten Tabelle gibt die Werte für die Risikoanfälligkeit an.
- **site.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Unternehmens geht, neue Standorte für die betriebliche Expansion auszuwählen. Das Unternehmen beauftragte zwei Berater unabhängig voneinander mit der Bewertung der Standorte. Neben einem umfassenden Bericht gaben die Berater auch eine zusammenfassende Wertung für jeden Standort als “good” (gut) “fair” (mittelmäßig) oder “poor” (schlecht) ab.

- **siteratings.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Betatests der neuen Website eines E-Commerce-Unternehmens geht. Jeder Fall entspricht einem Beta-Tester, der die Brauchbarkeit der Website auf einer Skala von 0 bis 20 bewertete.
- **smokers.sav.** Diese Datendatei wurde aus der Umfrage "National Household Survey of Drug Abuse" aus dem Jahr 1998 abstrahiert und stellt eine Wahrscheinlichkeitsstichprobe US-amerikanischer Haushalte dar. Daher sollte der erste Schritt bei der Analyse dieser Datendatei darin bestehen, die Daten entsprechend den Bevölkerungstrends zu gewichten.
- **smoking.sav.** Hierbei handelt es sich um eine von Greenacre (Greenacre , 1984) vorgestellte hypothetische Tabelle. Die relevante Tabelle wird durch eine Kreuztabelle der Rauchgewohnheiten und der Berufskategorie gebildet. Die Variable *Berufsgruppe* enthält die Berufskategorien *Senior Manager*, *Junior Manager*, *Angestellter mit Erfahrung*, *Angestellter ohne Erfahrung* und *Sekretariat* sowie die Kategorie *National Average*, die als Ergänzung der Analyse dienen kann. Die Variable *Rauchen* enthält die Rauchgewohnheiten *Nichtraucher*, *Leicht*, *Mittel* und *Stark* sowie die Kategorien *No Alcohol* und *Alcohol*, die als Ergänzung der Analyse dienen können.
- **storebrand.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Verkaufsleiterin in einem Lebensmittelmarkt geht, die die Verkaufszahlen des Waschmittels der Eigenmarke gegenüber den anderen Marken steigern möchte. Sie erarbeitet eine Werbeaktion im Geschäft und spricht an der Kasse mit Kunden. Jeder Fall entspricht einem Kunden.
- **stores.sav.** Diese Datendatei enthält hypothetische monatliche Marktanteilsdaten für zwei konkurrierende Lebensmittelgeschäfte. Jeder Fall entspricht den Marktanteilsdaten für einen bestimmten Monat.
- **stroke_clean.sav.** Diese hypothetische Datendatei enthält den Zustand einer medizinischen Datenbank, nachdem diese mithilfe der Prozeduren in der Option "Data Preparation" bereinigt wurde.
- **stroke_invalid.sav.** Diese hypothetische Datendatei enthält den ursprünglichen Zustand einer medizinischen Datenbank, der mehrere Dateneingabefehler aufweist.
- **stroke_survival.** In dieser hypothetischen Datendatei geht es um die Überlebenszeiten von Patienten, die nach einem Rehabilitationsprogramm wegen eines ischämischen Schlaganfalls mit einer Reihe von Problemen zu kämpfen haben. Nach dem Schlaganfall werden das Auftreten von Herzinfarkt, ischämischem Schlaganfall und hämorrhagischem Schlaganfall sowie der Zeitpunkt des Ereignisses aufgezeichnet. Die Stichprobe ist auf der linken Seite abgeschnitten, da sie nur Patienten enthält, die bis zum Ende des Rehabilitationprogramms, das nach dem Schlaganfall durchgeführt wurde, überlebten.
- **stroke_valid.sav.** Diese hypothetische Datendatei enthält den Zustand einer medizinischen Datenbank, nachdem diese mithilfe der Prozedur "Daten validieren" überprüft wurde. Sie enthält immer noch potenziell anomale Fälle.
- **tastetest.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bewertung der Auswirkungen der Mulchfarbe auf den Geschmack von Pflanzenprodukten geht. Der Geschmack von Erdbeeren, die in rotem, blauem und schwarzem Rindenmulch gezogen wurden, wurde von Testpersonen auf einer ordinalen Skala (weit unter bis weit über dem Durchschnitt) bewertet. Jeder Fall entspricht einem Geschmackstester.

- **telco.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Telekommunikationsunternehmens geht, die Kundenabwanderung zu verringern. Jeder Fall entspricht einem Kunden und enthält verschiedene Informationen zum demografischen Hintergrund und zur Servicenutzung.
- **telco_extra.sav.** Diese Datendatei ähnelt der Datei *telco.sav*, allerdings wurden die Variablen "tenure" und die Log-transformierten Variablen zu den Kundenausgaben entfernt und durch standardisierte Log-transformierte Variablen ersetzt.
- **telco_missing.sav.** Diese Datendatei entspricht der Datei *telco_mva_complete.sav*, allerdings wurde ein Teil der Daten durch fehlende Werte ersetzt.
- **telco_mva_complete.sav.** Bei dieser Datendatei handelt es sich um eine Teilmenge der Datendatei *telco.sav*, allerdings mit anderen Variablennamen.
- **testmarket.sav.** Diese hypothetische Datendatei bezieht sich auf die Pläne einer Fast-Food-Kette, einen neuen Artikel in ihr Menü aufzunehmen. Es gibt drei mögliche Kampagnen zur Verkaufsförderung für das neue Produkt. Daher wird der neue Artikel in Filialen in mehreren zufällig ausgewählten Märkten eingeführt. An jedem Standort wird eine andere Form der Verkaufsförderung verwendet und die wöchentlichen Verkaufszahlen für das neue Produkt werden für die ersten vier Wochen aufgezeichnet. Jeder Fall entspricht einer Standort-Woche.
- **testmarket_1month.sav.** Bei dieser hypothetischen Datendatei handelt es sich um die Datendatei *testmarket.sav*, wobei die wöchentlichen Verkaufszahlen zusammengefasst sind, sodass jeder Fall einem Standort entspricht. Dadurch entfallen einige der Variablen, die wöchentlichen Änderungen unterworfen waren, und die verzeichneten Verkaufszahlen sind nun die Summe der Verkaufszahlen während der vier Wochen der Studie.
- **tree_car.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zum Kaufpreis von Fahrzeugen enthält.
- **tree_credit.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zu früheren Bankkrediten enthält.
- **tree_missing_data.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zu früheren Bankkrediten enthält und eine große Anzahl fehlender Werte aufweist.
- **tree_score_car.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zum Kaufpreis von Fahrzeugen enthält.
- **tree_textdata.sav.** Eine einfache Datendatei mit nur zwei Variablen, die vor allem den Standardzustand von Variablen vor der Zuweisung von Messniveau und Wertelabels zeigen soll.
- **tv-survey.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einer Studie, die von einem Fernsehstudio durchgeführt wurde, das überlegt, ob die Laufzeit eines erfolgreichen Programms verlängert werden soll. 906 Personen wurden gefragt, ob sie das Programm unter verschiedenen Bedingungen ansehen würden. Jede Zeile entspricht einem Befragten; jede Spalte entspricht einer Bedingung.
- **ulcer_recurrence.sav.** Diese Datei enthält Teilinformationen aus einer Studie zum Vergleich der Wirksamkeit zweier Therapien zur Vermeidung des Wiederauftretens von Geschwüren. Es stellt ein gutes Beispiel für intervallzensierte Daten dar und wurde an anderer Stelle (Collett, 2003) vorgestellt und analysiert.

- **ulcer_recurrence_recoded.sav.** In dieser Datei sind die Daten aus *ulcer_recurrence.sav* so umstrukturiert, dass das Modell der Ereigniswahrscheinlichkeit für jedes Intervall der Studie berechnet werden kann und nicht nur die Ereigniswahrscheinlichkeit am Ende der Studie. Sie wurde an anderer Stelle (Collett et al., 2003) vorgestellt und analysiert.
- **verd1985.sav.** Diese Datendatei enthält eine Umfrage (Verdegaal, 1985). Die Antworten von 15 Subjekten auf 8 Variablen wurden aufgezeichnet. Die relevanten Variablen sind in drei Sets unterteilt. Set 1 umfasst *alter* und *heirat*, Set 2 besteht aus *pet* und *news* und in Set 3 finden sich *music* und *live*. Die Variable *pet* wird mehrfach nominal skaliert und die Variable *Alter* ordinal. Alle anderen Variablen werden einzeln nominal skaliert.
- **virus.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Internet-Dienstanbieters geht, der die Auswirkungen eines Virus auf seine Netzwerke ermitteln möchte. Dabei wurde vom Moment der Virusentdeckung bis zu dem Zeitpunkt, zu dem die Virusinfektion unter Kontrolle war, der (ungefähre) prozentuale Anteil infizierter E-Mail in den Netzwerken erfasst.
- **waittimes.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu den Wartezeiten für Kunden bei drei verschiedenen Filialen einer Bank. Jeder Fall entspricht einem Kunden und zeichnet die Wartezeit und die Filiale.
- **webusability.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Tests zur Benutzerfreundlichkeit eines neuen Internetgeschäfts geht. Jeder Fall entspricht einer von fünf Testpersonen, die die Benutzerfreundlichkeit bewerten und gibt für sechs separate Aufgaben an, ob die Testperson sie erfolgreich ausführen könnte.
- **wheeze_steubenville.sav.** Hierbei handelt es sich um eine Teilmenge der Daten aus einer Langzeitstudie zu den gesundheitlichen Auswirkungen der Luftverschmutzung auf Kinder (Ware, Dockery, Spiro III, Speizer, als auch Ferris Jr., 1984). Die Daten enthalten wiederholte binäre Messungen des Keuchens von Kindern aus Steubenville, Ohio, im Alter von 7, 8, 9 und 10 Jahren sowie eine unveränderlichen Angabe, ob die Mutter im ersten Jahr der Studie rauchte oder nicht.
- **workprog.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einem Arbeitsprogramm der Regierung, das versucht, benachteiligten Personen bessere Arbeitsplätze zu verschaffen. Eine Stichprobe potenzieller Programmteilnehmer wurde beobachtet. Von diesen Personen wurden nach dem Zufallsprinzip einige für die Teilnahme an dem Programm ausgewählt. Jeder Fall entspricht einem Programmteilnehmer.

Bibliografie

Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.

Blake, C. L., als auch C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at http://www.ics.uci.edu/~mlearn/MLRepository.html.

Breiman, L., als auch J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580–598.

Cochran, W. G. 1977. *Sampling Techniques*, 3rd (Hg.). New York: John Wiley and Sons.

Collett, D. 2003. *Modelling survival data in medical research*, 2 (Hg.). Boca Raton: Chapman & Hall/CRC.

Cox, D. R., als auch E. J. Snell. 1989. *The Analysis of Binary Data*, 2nd (Hg.). London: Chapman and Hall.

Green, P. E., als auch V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

Green, P. E., als auch Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

Greenacre , M. J. 1984. *Theory and applications of correspondence analysis*. London: Academic Press.

Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, 469–506.

Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.

Hastie, T., als auch R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.

Kennedy, R., C. Riquier, als auch B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, 56–70.

Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons.

Kish, L. 1987. *Statistical Design for Research*. New York: John Wiley and Sons.

McCullagh, P., als auch J. A. Nelder. 1989. *Generalized Linear Models*, 2nd (Hg.). London: Chapman & Hall.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In: *Frontiers in Economics,* P. Zarembka (Hg.). New York: Academic Press.

Menec , V., N. Roos, D. Nowicki, L. MacWilliam, G. Finlayson , als auch C. Black. 1999. *Seasonal Patterns of Winnipeg Hospital Use*. : Manitoba Centre for Health Policy.

Murthy, M. N. 1967. *Sampling Theory and Methods*. Kalkutta, Indien: Statistical Publishing Society.

Nagelkerke, N. J. D. 1991. A note on the general definition of the coefficient of determination. *Biometrika*, 78:3, 691–692.

Price, R. H., als auch D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, 579–586.

Rickman, R., N. Mitchell, J. Dingman, als auch J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228, 54–58.

Rosenberg, S., als auch M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489–502.

Särndal, C., B. Swensson, als auch J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Van der Ham, T., J. J. Meulman, D. C. Van Strien, als auch H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363–368.

Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (in niederländischer Sprache)*. Leiden: Department of Data Theory, Universität Leiden.

Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, als auch B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, 366–374.

Index

Abweichungskontraste
  im allgemeinen linearen Modell für komplexe Stichproben, 52
Abweichungsresiduen
  in Cox-Regression für komplexe Stichproben, 92
Aggregierte Residuen
  in Cox-Regression für komplexe Stichproben, 92
Allgemeines lineares Modell für komplexe Stichproben, 46, 186
  geschätzte Mittelwerte, 52
  Modell, 48
  Modellzusammenfassung, 191
  Optionen, 54
  Parameterschätzer, 192
  Randmittel, 193
  Statistiken, 49
  Tests der Modelleffekte, 191
  Variablen speichern, 53
  verwandte Prozeduren, 196
  zusätzliche Funktionen beim Befehl, 54
Analyseplan, 19
Analysevorbereitungsassistent für komplexe Stichproben, 148
  Auswertung, 151, 162
  keine Stichprobengewichte verfügbar, 151
  öffentliche Daten, 148
  verwandte Prozeduren, 162
Antwortwahrscheinlichkeiten
  bei der ordinalen Regression für komplexe Stichproben, 68
Auswertung
  im Analysevorbereitungsassistenten, 151, 162

Basisschichten
  in Cox-Regression für komplexe Stichproben, 85
Beispieldateien
  Speicherort, 271
Bonferroni
  in Cox-Regression für komplexe Stichproben, 91
  in Komplexe Stichproben, 50, 61, 72
Breslow-Schätzmethode
  in Cox-Regression für komplexe Stichproben, 96
Brewers Stichprobenmethode
  beim Stichprobenassistenten, 8

Chi-Quadrat
  in Cox-Regression für komplexe Stichproben, 91
  in Komplexe Stichproben, 50, 61, 72
Cox-Regression für komplexe Stichproben, 223
  Datums- und Uhrzeit-Variablen, 78
  Diagramme, 90
  Einflussvariablen, 82
  Ereignis definieren, 81
  Hypothesentests, 91
  Informationen zum Stichprobenplan, 233, 266
  Kaplan-Meier-Analyse, 78
  Log-Minus-Log-Diagramm, 269
  Modell, 86
  Modellexport, 94
  Musterwerte, 268
  Optionen, 96
  Parameterschätzer, 238, 267
  Statistik, 88
  stückweise konstante, zeitabhängige Einflussvariablen, 238
  Test für proportionale Hazard-Raten, 234
  Tests der Modelleffekte, 234, 237, 267
  Untergruppen, 85
  Variablen speichern, 92
  Zeitabhängige Einflussvariable, 83, 223
Cox-Snell-Residuen
  in Cox-Regression für komplexe Stichproben, 92

Deskriptive Statistiken für komplexe Stichproben, 33, 169
  Fehlende Werte, 35
  öffentliche Daten, 169
  Statistiken, 34, 172
  Statistiken nach Teilgesamtheit, 172
  verwandte Prozeduren, 173
Differenzkontraste
  im allgemeinen linearen Modell für komplexe Stichproben, 52

Effekt des Stichprobenplans
  bei der logistischen Regression für komplexe Stichproben, 60
  bei der ordinalen Regression für komplexe Stichproben, 70
  im allgemeinen linearen Modell für komplexe Stichproben, 49
  in Cox-Regression für komplexe Stichproben, 88
  in Deskriptive Statistiken für komplexe Stichproben, 34
  in Häufigkeiten für komplexe Stichproben, 30
  in Kreuztabellen für komplexe Stichproben, 39
  in Verhältnisse für komplexe Stichproben, 43
Efron-Schätzmethode
  in Cox-Regression für komplexe Stichproben, 96
Einfache Kontraste
  im allgemeinen linearen Modell für komplexe Stichproben, 52
Einfache Zufallsstichprobenziehung
  beim Stichprobenassistenten, 8

Einflussvariablenmuster
in Cox-Regression für komplexe Stichproben, 268
Eingabe-Stichprobengewichtung
beim Stichprobenassistenten, 6
Einschlusswahrscheinlichkeiten
beim Stichprobenassistenten, 12
erwartete Werte
in Kreuztabellen für komplexe Stichproben, 39

F (korrigiert), Statistik
in Cox-Regression für komplexe Stichproben, 91
in Komplexe Stichproben, 50, 61, 72
F-Statistik
in Cox-Regression für komplexe Stichproben, 91
in Komplexe Stichproben, 50, 61, 72
Fehlende Werte
bei der logistischen Regression für komplexe Stichproben, 64
bei der ordinalen Regression für komplexe Stichproben, 75
im allgemeinen linearen Modell für komplexe Stichproben, 54
in Deskriptive Statistiken für komplexe Stichproben, 35
in Komplexe Stichproben, 31, 40
in Verhältnisse für komplexe Stichproben, 44
Fisher-Bewertung
bei der ordinalen Regression für komplexe Stichproben, 75
Freiheitsgrade
in Cox-Regression für komplexe Stichproben, 91
Freiheitsgrade.
in Komplexe Stichproben, 50, 61, 72

Geringste signifikante Differenz
in Cox-Regression für komplexe Stichproben, 91
in Komplexe Stichproben, 50, 61, 72
Geschätzte Randmittel
im allgemeinen linearen Modell für komplexe Stichproben, 52

Häufigkeiten für komplexe Stichproben, 29, 163
Häufigkeitstabelle, 166
Häufigkeitstabelle nach Teilgesamtheit, 167
Statistiken, 30
verwandte Prozeduren, 168
Helmert-Kontraste
im allgemeinen linearen Modell für komplexe Stichproben, 52

Informationen zum Stichprobenplan
in Cox-Regression für komplexe Stichproben, 88, 233, 266
Iterationen
bei der logistischen Regression für komplexe Stichproben, 64
bei der ordinalen Regression für komplexe Stichproben, 75
Iterationsprotokoll
bei der logistischen Regression für komplexe Stichproben, 64
bei der ordinalen Regression für komplexe Stichproben, 75

Klassifikationstabellen
bei der logistischen Regression für komplexe Stichproben, 60, 202
bei der ordinalen Regression für komplexe Stichproben, 70, 215
Klumpen
beim Stichprobenassistenten, 6
im Analysevorbereitungsassistenten, 21
komplexe Stichproben
Analyseplan, 19
Stichprobenplan, 4
Komplexe Stichproben
Fehlende Werte, 31, 40
Hypothesentests, 50, 61, 72
Optionen, 32, 36, 41, 45
Konfidenzintervalle
bei der logistischen Regression für komplexe Stichproben, 60
bei der ordinalen Regression für komplexe Stichproben, 70
im allgemeinen linearen Modell für komplexe Stichproben, 49, 54
in Deskriptive Statistiken für komplexe Stichproben, 34, 172
in Häufigkeiten für komplexe Stichproben, 30, 166–167
in Kreuztabellen für komplexe Stichproben, 39
in Verhältnisse für komplexe Stichproben, 43
Konfidenzniveau
bei der logistischen Regression für komplexe Stichproben, 64
bei der ordinalen Regression für komplexe Stichproben, 75
Kontraste
im allgemeinen linearen Modell für komplexe Stichproben, 52
Korrelationen der Parameterschätzer
bei der logistischen Regression für komplexe Stichproben, 60
bei der ordinalen Regression für komplexe Stichproben, 70
im allgemeinen linearen Modell für komplexe Stichproben, 49
Korrigierte Residuen
in Kreuztabellen für komplexe Stichproben, 39
Korrigiertes Chi-Quadrat
in Cox-Regression für komplexe Stichproben, 91
in Komplexe Stichproben, 50, 61, 72

Kovarianzen der Parameterschätzer
bei der logistischen Regression für komplexe Stichproben, 60
bei der ordinalen Regression für komplexe Stichproben, 70
im allgemeinen linearen Modell für komplexe Stichproben, 49
Kreuztabellen für komplexe Stichproben, 37, 174
Relatives Risiko, 174, 178–179
Statistiken, 39
Tabelle "Kreuztabelle", 177
verwandte Prozeduren, 180
Kumulative Wahrscheinlichkeiten
bei der ordinalen Regression für komplexe Stichproben, 74
Kumulative Werte
in Häufigkeiten für komplexe Stichproben, 30

Likelihood-Konvergenz
bei der logistischen Regression für komplexe Stichproben, 64
bei der ordinalen Regression für komplexe Stichproben, 75
Log-Minus-Log-Diagramm
in Cox-Regression für komplexe Stichproben, 269
Logistische Regression für komplexe Stichproben, 56, 197
Klassifikationstabellen, 202
Modell, 58
Optionen, 64
Parameterschätzer, 203
Pseudo-R^2-Statistik, 201
Quotenverhältnis, 62, 204
Referenzkategorie, 57
Statistiken, 60
Tests der Modelleffekte, 203
Variablen speichern, 63
verwandte Prozeduren, 206
zusätzliche Funktionen beim Befehl, 65

Martingal-Residuen
in Cox-Regression für komplexe Stichproben, 92
Maß für die Größe der Einheiten
beim Stichprobenassistenten, 8
Methode der Stichprobenziehung
beim Stichprobenassistenten, 8
Mittelwert
in Deskriptive Statistiken für komplexe Stichproben, 34, 172
Murthys Stichprobenmethode
beim Stichprobenassistenten, 8

Newton-Raphson-Methode
bei der ordinalen Regression für komplexe Stichproben, 75

öffentliche Daten
im Analysevorbereitungsassistenten, 148
in Deskriptive Statistiken für komplexe Stichproben, 169
Ordinale Regression für komplexe Stichproben, 66, 207
Antwortwahrscheinlichkeiten, 68
Klassifikationstabellen, 215
Modell, 69
Optionen, 75
Parameterschätzer, 213
Pseudo-R^2-Statistik, 212, 221
Quotenverhältnis, 73, 216
Statistiken, 70
Tests der Modelleffekte, 213
Variablen speichern, 74
Verallgemeinertes kumulatives Modell, 217
verwandte Prozeduren, 222
Warnungen, 220

Parallelitätstest für Linien
bei der ordinalen Regression für komplexe Stichproben, 70, 217
Parameter-Konvergenz
bei der logistischen Regression für komplexe Stichproben, 64
bei der ordinalen Regression für komplexe Stichproben, 75
Parameterschätzer
bei der logistischen Regression für komplexe Stichproben, 60, 203
bei der ordinalen Regression für komplexe Stichproben, 70, 213
im allgemeinen linearen Modell für komplexe Stichproben, 49, 192
in Cox-Regression für komplexe Stichproben, 88
Partielle Schoenfeld-Residuum
in Cox-Regression für komplexe Stichproben, 92
Plandatei, 3
Polynomiale Kontraste
im allgemeinen linearen Modell für komplexe Stichproben, 52
PPS-Stichprobenziehung
beim Stichprobenassistenten, 8
Pseudo-R^2-Statistik
bei der logistischen Regression für komplexe Stichproben, 60, 201
bei der ordinalen Regression für komplexe Stichproben, 70, 212, 221

Quadratwurzel aus dem Effekt des Stichprobenplans
bei der logistischen Regression für komplexe Stichproben, 60
bei der ordinalen Regression für komplexe Stichproben, 70
im allgemeinen linearen Modell für komplexe Stichproben, 49
in Cox-Regression für komplexe Stichproben, 88

in Deskriptive Statistiken für komplexe Stichproben, 34
in Häufigkeiten für komplexe Stichproben, 30
in Kreuztabellen für komplexe Stichproben, 39
in Verhältnisse für komplexe Stichproben, 43
Quotenverhältnis
bei der logistischen Regression für komplexe Stichproben, 62, 204
bei der ordinalen Regression für komplexe Stichproben, 73, 216
in Kreuztabellen für komplexe Stichproben, 39, 174

R^2-Statistik
im allgemeinen linearen Modell für komplexe Stichproben, 49, 191
Randmittel
in GLM - Univariat, 193
Referenzkategorie
bei der logistischen Regression für komplexe Stichproben, 57
im allgemeinen linearen Modell für komplexe Stichproben, 52
Relatives Risiko
in Kreuztabellen für komplexe Stichproben, 39, 174, 178–179
Residuen
im allgemeinen linearen Modell für komplexe Stichproben, 53
in Kreuztabellen für komplexe Stichproben, 39
Risiko-Differenz
in Kreuztabellen für komplexe Stichproben, 39

Sampfords Stichprobenmethode
beim Stichprobenassistenten, 8
Schichtung
beim Stichprobenassistenten, 6
im Analysevorbereitungsassistenten, 21
Schritt-Halbierungen
bei der logistischen Regression für komplexe Stichproben, 64
bei der ordinalen Regression für komplexe Stichproben, 75
Sequenzielle Bonferroni-Korrektur
in Cox-Regression für komplexe Stichproben, 91
in Komplexe Stichproben, 50, 61, 72
Sequenzielle Sidak-Korrektur
in Cox-Regression für komplexe Stichproben, 91
in Komplexe Stichproben, 50, 61, 72
Sequenzielle Stichprobenziehung
beim Stichprobenassistenten, 8
Sidak-Korrektur
in Cox-Regression für komplexe Stichproben, 91
in Komplexe Stichproben, 50, 61, 72
Spaltenprozente
in Kreuztabellen für komplexe Stichproben, 39
Standardfehler
bei der logistischen Regression für komplexe Stichproben, 60
bei der ordinalen Regression für komplexe Stichproben, 70
im allgemeinen linearen Modell für komplexe Stichproben, 49
in Deskriptive Statistiken für komplexe Stichproben, 34, 172
in Häufigkeiten für komplexe Stichproben, 30, 166–167
in Kreuztabellen für komplexe Stichproben, 39
in Verhältnisse für komplexe Stichproben, 43
Stichprobe
komplexer Plan, 4
Stichprobenanteil
beim Stichprobenassistenten, 12
Stichprobenassistent für komplexe Stichproben, 100
PPS-Stichprobenziehung, 130
Stichprobenrahmen, partiell, 112
Stichprobenrahmen, vollständig, 100
verwandte Prozeduren, 147
Zusammenfassung, 110, 142–143
Stichprobengewichtungen
beim Stichprobenassistenten, 12
im Analysevorbereitungsassistenten, 21
Stichprobenplan, 4
Stichprobenrahmen, partiell
beim Stichprobenassistenten, 112
Stichprobenrahmen, vollständig
beim Stichprobenassistenten, 100
Stichprobenschätzung
im Analysevorbereitungsassistenten, 22
Stichprobenumfang
beim Stichprobenassistenten, 10, 12
stückweise konstante, zeitabhängige Einflussvariablen
in Cox-Regression für komplexe Stichproben, 238
Summe
in Deskriptive Statistiken für komplexe Stichproben, 34
Systematische Stichprobenziehung
beim Stichprobenassistenten, 8

T-Test
bei der logistischen Regression für komplexe Stichproben, 60
bei der ordinalen Regression für komplexe Stichproben, 70
im allgemeinen linearen Modell für komplexe Stichproben, 49
Tabelle “Kreuztabelle”
in Kreuztabellen für komplexe Stichproben, 177
Tabellenprozente
in Häufigkeiten für komplexe Stichproben, 30, 166–167
in Kreuztabellen für komplexe Stichproben, 39
Teilgesamtheit
in Cox-Regression für komplexe Stichproben, 85
Test für proportionale Hazard-Raten
in Cox-Regression für komplexe Stichproben, 234
Test proportionaler Hazards
in Cox-Regression für komplexe Stichproben, 88

Tests der Modelleffekte
bei der logistischen Regression für komplexe Stichproben, 203
bei der ordinalen Regression für komplexe Stichproben, 213
im allgemeinen linearen Modell für komplexe Stichproben, 191
in Cox-Regression für komplexe Stichproben, 267
Trennung
bei der logistischen Regression für komplexe Stichproben, 64
bei der ordinalen Regression für komplexe Stichproben, 75

Umfang der Grundgesamtheit
beim Stichprobenassistenten, 12
in Deskriptive Statistiken für komplexe Stichproben, 34
in Häufigkeiten für komplexe Stichproben, 30, 166–167
in Kreuztabellen für komplexe Stichproben, 39
in Verhältnisse für komplexe Stichproben, 43
Ungewichtete Anzahl
in Deskriptive Statistiken für komplexe Stichproben, 34
in Häufigkeiten für komplexe Stichproben, 30
in Kreuztabellen für komplexe Stichproben, 39
in Verhältnisse für komplexe Stichproben, 43

Variationskoeffizient (COV)
in Deskriptive Statistiken für komplexe Stichproben, 34
in Häufigkeiten für komplexe Stichproben, 30
in Kreuztabellen für komplexe Stichproben, 39
in Verhältnisse für komplexe Stichproben, 43
Verallgemeinertes kumulatives Modell
bei der ordinalen Regression für komplexe Stichproben, 217
Verhältnisse
in Verhältnisse für komplexe Stichproben, 183
Verhältnisse für komplexe Stichproben, 42, 181
Fehlende Werte, 44
Statistiken, 43
Verhältnisse, 183
verwandte Prozeduren, 185
Vorhergesagte Kategorien
bei der logistischen Regression für komplexe Stichproben, 63
bei der ordinalen Regression für komplexe Stichproben, 74
Vorhergesagte Wahrscheinlichkeit
bei der logistischen Regression für komplexe Stichproben, 63
bei der ordinalen Regression für komplexe Stichproben, 74
Vorhergesagte Werte
im allgemeinen linearen Modell für komplexe Stichproben, 53

Warnungen
bei der ordinalen Regression für komplexe Stichproben, 220
Wertresiduen
in Cox-Regression für komplexe Stichproben, 92
Wiederholte Kontraste
im allgemeinen linearen Modell für komplexe Stichproben, 52

Zeilenprozente
in Kreuztabellen für komplexe Stichproben, 39
Zeitabhängige Einflussvariable
in Cox-Regression für komplexe Stichproben, 83, 223
Zusammenfassung
beim Stichprobenassistenten, 110, 142–143